# Charge Engineering of Alternative Protein Scaffolds for Molecular Imaging

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA

BY

Brett Alynn Case

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Advisor: Benjamin J. Hackel

June 2018

# Abstract

Molecular imaging is a powerful, noninvasive technique for patient stratification and treatment monitoring via cancer biomarker characterization with reporter probes. Small scaffold proteins that combine stable frameworks with diversifiable paratope regions are well-suited as engineerable probes for molecular imaging and often benefit from efficient physiological transport. Probe efficacy depends not only on diseased region localization but also passive noninteraction with plasma molecules and healthy tissues. Precise relationships between protein charge, physicochemical robustness, and *in vivo* performance are poorly understood, but the appropriate manipulation of ionic charge can stabilize native protein folds and reduce nonspecific interactions. Therefore, the effective engineering of protein charge in diagnostic molecular probes would advance cancer characterization and personalized medicine.

In this thesis work we enlisted affibody and Gp2 proteins as molecular imaging scaffolds, applied discrete protein engineering designs to modify charge density and distribution, and evaluated promising charge-modified variants as positron emission tomography (PET) imaging agents in preclinical studies against the cancer biomarker, epidermal growth factor receptor (EGFR). Natural and functional (synthetic) homolog frequencies guided the design of charge-modified consensus clones and combinatorial libraries based on a hybrid anti-EGFR affibody, EA68; a synthetic consensus design proved most robust by providing variants that retain parental recombinant yield and stability while improving EGFR affinity. Charge-mutated combinatorial Gp2 libraries were enriched for binding against three clinically or biotechnologically relevant targets and revealed target-

dependent total, net, and site-specific charge preferences demonstrating the utility of multitarget screening to determine globally accepted framework mutations. In summary, hybrids of combinatorial library selection and sequence informatics-driven design have proven effective for charge engineering.

Preclinical investigation of lead anti-EGFR affibody and Gp2 clones and their charge-varied homologs in mice bearing $EGFR^{high}$ and $EGFR^{low}$ tumor xenografts exposed differences in clonal biodistribution, tumor targeting, and proteolytic stability. Among tested variants, the charge-reduced affibody 'EA35S' and Gp2 'GαE35' achieved the best overall performance with high (12 ± 2 %ID/g and 5.5 ± 0.5%ID/g) and specific (tumor:muscle = 34 ± 5 and 35 ± 5; tumor:blood = 12 ± 3 and 9.7 ± 1.3) $EGFR^{high}$ tumor localization at 4 and 2 h post-injection.

Collectively, this work provides a suite of tools to maintain and enhance protein function through ionic charge engineering and reveals two promising probes for cancer characterization through molecular recognition of EGFR.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1 - Introduction

## 1.1. Proteins provide diverse functionality

Proteins are a versatile class of molecules that have proven effective in industrial and biotechnological catalysis, targeted clinical therapy and diagnostics, and other applications. Degradative enzymes including amylases, lipases, phytases, and proteases are widely used in food[1], detergent[2], textile[3], animal feed[4], and biofuel industries[5]. To improve or alter their catalytic function, protein engineering has been used to enhance thermal stability[6–8], tolerance to acidic[9] and basic[5] conditions, and stereoselectivity[10]. Additionally, numerous proteins are employed in fundamental biotechnological applications such as fluorescent reporter molecules[11,12], polymerases[13] and ligases[14] for DNA manipulation, and peroxidases and phosphatases in antigen detection assays[15].

The utilization of proteins in medical research, therapy, and diagnostics is substantial. Peptide-based biopolymers including elastin-like polypeptides, leucine zipper hydrogels, and silk-like polymers have been designed to deliver cytotoxic loads under localized hyperthermia near tumors[16], retard the *in vivo* release of therapeutics [17], and form self-assembling hydrogels as scaffold materials for tissue engineering[18,19]. Antimicrobial, antiviral, and cell-penetrating peptides are used to selectively combat gram-positive[20] and gram-negative[21,22] bacterial pathogens, inhibit viral infections[23,24], and intracellularly transport therapeutic cargo[25,26]. Antibodies have been engineered to inhibit native receptor-ligand interactions[27,28], bridge tumor cells and immune cells to enhance interaction[29], and deliver drugs to receptor-expressing tumor cells[30] in monovalent, bispecific, and protein-drug conjugate formats, respectively. Over 200 monoclonal and polyclonal antibodies are

available for use in immunohistochemical staining assays to detect cancer cell cycle, angiogenic, apoptotic, and patient stratifying antigens.[31–34] As such, an understanding of the protein sequence/function relationship to enable the engineering of proteins for such applications is an important area for scientific and technological development.

### 1.2. Protein engineering must balance numerous functions and properties

The utility of a protein is dependent upon not only its core function of molecular binding, enzymatic catalysis, or material strength but also upon the physicochemical robustness of the protein and its ability to passively lack interaction with nontarget, nonsubstrate molecules in its environment. Numerous auxiliary characteristics influence protein performance including structural stability[35], size[36], solubility[37], hydrophobicity[38], ionic charge[39], target discrimination[40], and physiological transport kinetics[41,42] with considerable interplay between each of these attributes. Engineering of useful proteins with enhanced or novel function requires consideration of these factors.

Retention of native structure is a common prerequisite for proper protein function[43], but the phenotypical diversity required for novel or enhanced functionality may necessitate mutations that destabilize the protein fold[44]. Antecedent thermodynamic stabilization promotes efficient discovery by increasing mutational tolerance.[45] Additionally, thermal and chemical stability provide resistance to denaturation during stringent biotechnological processes, maintaining functional potency and allowing for repeated usage. Improving proteolytic resistance, which often correlates to increased thermal stability[46], is critical for enzymes that are mixed with proteases during industrial applications[47]. Limiting the formation of insoluble multimers (inclusion bodies) during prokaryotic expression

increases recombinant protein yield, mitigating the need for expensive denaturation and refolding procedures.[48] Furthermore, protein aggregates may be toxic[49] *in vivo* and are associated with an elevated immune response[50]. Both the stabilization of native structures[51,52] and modification of "sticky" hydrophobic patches in well-folded proteins[53] are valuable methods by which to hinder aggregation.

Prevention of nonsubstrate and nontarget interactions reduces undesirable enzymatic byproducts, improves drug therapeutic indices, and increases contrast for disease state imaging. While specific off-target interactions may be reduced with advanced knowledge of interfacial residues[40], physicochemical properties such as size, surface area, and hydrophobicity are often better suited to address broad nonspecificity. Proteins administered for therapeutic and detection purposes, such as antibodies and antibody-derived fragments, often possess hydrophilic surfaces which cause them to tarry in blood and interstitial space before excretion. As such, scaffolds with molecular weights below the glomerular filtration cut-off (~ 50 kDa) undergo rapid plasma clearance[36] thereby reducing instances of undesirable nontarget interactions. Additionally, small proteins and peptides bear less surface area, reducing the potential for unwanted protein interfaces. While the relationship between protein hydrophobicity and biodistribution is not well-understood and often scaffold dependent[54,55], hydrophobicity has been positively and negatively correlated with hepatic and renal signal, respectively[38] suggesting a method by which excretory organ retention may be altered.

### 1.3. Ionic charge impacts protein performance yet is difficult to engineer

The physicochemical property of primary interest in this thesis is ionic charge, which impacts performance by altering protein stability, solubility, and *in vivo* biodistribution and pharmacokinetic profile. Under aqueous conditions, the introduction of an ionizable functional group from an electrically charged amino acid into the hydrophobic interior of a folded protein is typically destabilizing.[56–58] As such, charged residues often reside in solvent exposed positions to impart thermodynamic stability and avoid the energetic penalty associated with desolvation.[59,60] However, internal acidic and basic residues are commonly required for enzymatic transduction pathways, which necessitates protein structure with high thermostability to retain functionality.[56] Ionizable residues also affect the relative strength of intramolecular hydrogen bonds. Charge-stabilized hydrogen bonds are, on average, ~2 kcal $mol^{-1}$ stronger than neutral hydrogen bonds.[61] Electrostatic attraction between oppositely charged functional groups contribute several kcal $mol^{-1}$ to scaffold stability in both long-distance coulombic interactions[62,63] and short-distance ($\leq$ 4 Å) ion pairs[64,65]. Solubility is also impacted by net charge. The isoelectric point, or pH at which a protein carries no net charge, is dependent upon acid dissociation contributions from charged amino acids within the molecule. In solutions with a pH near the protein's isoelectric point charges are balanced, causing a reduction in repulsive electrostatic forces and increased aggregation and precipitation.[66] Additionally, patches of positive surface charge have been implicated in reduced protein solubility during production in prokaryotic expression systems due to nucleic acid binding.[67]

Ionic charge significantly impacts protein performance *in vivo*, affecting ligand specificity, vascular permeability, tumor uptake, systemic clearance rates, and

biodistribution. Mammalian cell membranes are natively negative due, in part, to glycosylation of surface proteins and lipids. Highly positive proteins have been shown to interact nonspecifically with anionic cell membranes[68–71] causing undesirable nontarget localization. Furthermore, repellent electrostatic interactions between the ligand paratope[72,73] or nonparatope[74] regions with an antigen or epitope of interest can be thermodynamically unfavorable, hindering molecular recognition efforts. Cationic charge has been linked to an increase in solid tumor extravasation and retention for IgG[75], albumin[76], liposomes[77], and nanoparticles[78] as well as an elevation in plasma clearance[75]. However, highly anionic branched polypeptides also demonstrated rapid clearance suggesting an abstruse relationship between charge directionality and systemic retention.[39] Electric charge also impacts anatomical biodistribution. Highly cationic and anionic states have amplified renal retention for cystine knots[79], affibodies with chelator modifications[80,81], branched polypeptides[39], and bombesin analogs[82]. Conflictingly, reduced total protein charge was ineffective at modifying renal uptake in both hydrophilic and hydrophobic fibronectin domains, but a decrease in formal positive charge was associated with a drop in kidney retention.[38] Increases in both negative and positive charge through succinylation or amidation of polypeptides[39] or bovine serum albumin[83] resulted in elevated hepatic retention. Anti-Tac Fab fragments with increased positive or negative charge yield a two-fold increase in lung absorption or three-fold higher stomach localization, respectively.[84] Taken together, these outcomes suggest a complex relationship between protein structure; total, net, and distribution of charge; and *in vitro* and *in vivo* characteristics.

The ability to modulate charge to enhance physicochemical robustness and environmental passivity while retaining or enhancing function is a challenging, but achievable, goal. Naturally occurring proteins were often evolved for functionality rather than thermal stability[85,86] and interventional residue mutagenesis typically results in further destabilization[87,88]. To overcome stability penalties upon charge mutation while retaining or improving functionality, researchers have relied on protein engineering techniques including rational design, semi-rational design, and directed evolution.[89] Rational design and consensus modeling of charge-reversing or neutralizing mutations across surface residues has been successfully used to increase protein solubility[90,91] and stability[57,63]. Raghunathan et al. were able to explore surface charge modulation on the green fluorescent protein (GFP) without affecting its native fluorescence through co-introduction of known stabilizing mutations.[92] Neutral-to-charged residue mutations near the active site of *Pseudomonas glumae* lipase reduced its proteolytic susceptibility to substilin 5-fold compared to wild-type with minimal impact on catalytic activity.[47] Through directed evolution, Pédelacq et al. discovered two mutations, E40K and G33D, which increased the solubility of nucleoside diphosphate kinase 50-95% by improving charge distribution and hydrogen bonding.[93] Site-directed mutagenesis of a single, solvent-exposed position on ribonuclease Sa from tryptophan to aspartic or glutamic acid increased solubility nearly 12-fold near the proteins isoelectric point.[91] *In silico* stability analysis of the DNA-binding Sso7d protein by Traxlmyer et al. was used to reduce the net formal charge of Sso7d from +7 to +1 through basic residue neutralization and C-terminal lysine deletions while retaining parental clone hyperthermostability.[73]

Charge engineering is also an effective tool to alter protein physiological distribution and pharmacokinetics. The replacement of a single alanine with glutamic acid on the C-terminal linker region significantly increased tumor uptake in a bombesin peptide analog. However, adding two more glutamic acids to the linker abrogated this benefit and elevated renal retention.[82] Elimination of either arginine or glutamic acids from the mercaptoacetyl-based technetium chelator of an anti-HER2 affibody caused a substantial reduction in renal retention at 4 h post-injection.[80,81] Mutation of one or two arginine residues to glutamic acids in the 10th type III domain of human fibronectin did not decrease renal signal.[94] However, a more substantial charge removal lowered kidney signal independent of clonal hydrophobicity.[38] Replacement of 7 neutral positions with acidic residues in the variable domain of the LMB-2 immunotoxin reduced liver uptake, causing a 3-fold decrease in hepatic necrosis without altering CD25 binding affinity.[95] Modifying the hexa-histidine purification tag to alternate between histidine and glutamic acids '$(HE)_3$' reduced hepatic uptake nearly 6-fold in $Z_{HER2:342}$ affibodies[96] with a similar, but less pronounced effect seen in the ADAPT scaffold[97].

Although successfully implemented in the examples listed, engineering of charge remains difficult and can be particularly challenging in the context of a small protein[38,90] where mutations encompass a significant fraction of the protein structure[98]. Precise relationships between protein charge, physicochemical properties, and *in vivo* performance remain elusive. In this thesis, we apply protein charge engineering to molecular targeting ligands within the context of PET imaging of epidermal growth factor receptor in cancer.

### 1.4. Molecular recognition improves noninvasive cancer patient stratification

Cancer is the second leading cause of the death in the United States with approximately 1 in 3 Americans developing an invasive malignancy during their lifetime. The overall cancer death rate in the U.S. has decreased by 26% from its peak in 1991 due to an upsurge in smoking cessation, utilization of detection strategies such as mammograms and colonoscopies, and advances in oncological treatments.[99] Mortality rates among several common types including lung, colorectal, breast (female), and prostate cancers have seen a steady decline in recent years.[100] However, rates in liver, uterine, and brain malignancies have increased[101] and an estimated 1.7 million new cancer diagnoses are still expected in 2018 resulting in roughly 610,000 deaths.[99] Therefore, considerable effort is still necessary for the advancement of cancer diagnosis, prognosis, and treatment options.

Traditional oncological treatments, such as chemotherapy, were not designed to be molecularly specific, relying instead on the high tumoral growth rate found in several cancers[102] to exercise cytotoxicity, resulting in associated damage to healthy cells and limited efficacy for slow-growing and heterogeneous tumors. In recent years, precision medicine has begun to take advantage of advancements in the genetic, proteomic, and metabolic profiling of cancers to stratify disease status, determine the most efficacious treatment plan, and tailor medicinal doses appropriately.

The molecular characterization of cancer through biomarker expression is a powerful method of predicting therapeutic response and improving patient outcome.[103] Therapeutic benefit from paclitaxel in node-positive breast cancer patients was associated with human epidermal growth factor receptor 2 expression, regardless of estrogen receptor status.[104] Positive treatment response to bevacizumab in patients with advanced gastric

cancer correlated to high baseline plasma levels of vascular endothelial growth factor-A and low expression of neuropilin-1.[105] The addition of onartuzumab to erlotinib treatment significantly improved progression-free and overall survival rates in patients with non-small cell lung cancer that over-expressed the tyrosine-protein kinase Met, independent of gene expression.[106]

A biomarker that has garnered considerable clinical interest is the epidermal growth factor receptor (EGFR), also known as HER1 or ErbB1. EGFR is a transmembrane glycoprotein containing an extracellular receptor domain and intracellular cytoplasmic tyrosine kinase domain whose dysregulation has been implicated in tumor survival, growth, and metastasis.[107] Upon agonistic ligand binding or due to overexpression, EGFR homodimerization or intrafamilial heterodimerization with other ErbB receptors activates downstream events associated with oncogenic cell proliferation and apoptotic resistance including Ras-Raf-MEK-ERK, AKT-PI3K, and PLC-γ1-PKC pathways.[108] Elevated EGFR levels are common among the majority of solid tumors[107] with overexpression a strong prognostic indicator of poor patient outlook in head and neck, ovarian, cervical, esophageal, and bladder cancers[109]. EGFR expression levels and gene copy number are predictors of clinical response to receptor (cetuximab) and tyrosine kinase (gefitinib, icotinib) inhibiting treatments in primary and metastatic colorectal, non-small-cell lung, and esophageal cancers.[110–114] In patients with HER2-positive primary breast cancer EGFR surface overexpression, but not copy number, was associated with poor disease-free survival and predictive of response to trastuzumab therapy.[115] Therefore, molecular imaging of EGFR could aid in the stratification of responders versus nonresponders to targeted therapies across a range of cancer types. While several probes have undergone

preclinical trials[94,116,117], further development and discovery is necessary as there are no clinical EGFR imaging probes.

Frequently, upon diagnosis and to aid in personalized treatment planning, biopsies of the primary tumor site are extracted and histological analysis used to differentiate by cancer type through cell structural abnormalities, antigen presence, genetic mutation, and genomic expression levels. Although a gold standard in cancer characterization, biopsies suffer from several drawbacks. Biomarker expression levels can differ between primary and metastatic tumor sites within the same patient[118,119] leading to designed treatment plans with unpredictable efficacy towards satellite tumors. Furthermore, recent comprehensive genomic analyses has revealed the occurrence of intratumoral heterogeneity which may provide phenotypical diversity necessary to resist singular oncological treatments, suggesting a need for continued molecular characterization during treatment.[120,121] Unfortunately, repeated biopsies are often unfeasible due to financial and quality of life concerns. To address this need, considerable effort has been made into noninvasive diagnostic imaging.

Traditional diagnostic imaging lacks molecular specificity. Cross-sectional imaging techniques including magnetic resonance (MRI), computed tomography (CT), and ultrasonography (Ultrasound) rely on anatomical, physiological, or metabolic phenomena as contrast agents to distinguish between healthy and diseased states. In recent years, the field of molecular imaging has emerged to address the shortcomings in sensitivity and specificity of traditional imaging through the molecular-level targeting. Imaging probes are engineered to selectively bind disease state biomarkers that are upregulated on or near neoplasms and passively clear from background tissues. Combining ligands with reporter

agents such as microbubbles, fluorophores, and radioisotopes and utilizing imaging modalities including single-photon emission computed tomography (SPECT), positron emission tomography (PET), fluorescence molecular tomography (FMT), and ultrasonography the detection, location, classification, and mobility of various cancers can be determined.

A modality with significant potential for molecular imaging is PET. PET contrast agents are created through the addition of short half-life positron emitting radionuclides such as $^{68}$Ga ($t_{1/2}$ ~ 1.1 h), $^{18}$F (~ 1.8 h), and $^{64}$Cu (~ 12.7 h) with biomarker targeted ligands through covalent linkage[122] or radiometal chelating moiety[123]. As the radioisotope undergoes positive beta decay a positron is emitted, travelling a short distance[124–126] (~ 0.1 – 1 cm) before annihilating with an electron that causes the simultaneous discharge of two opposing γ-rays which are detected by surrounding scintillators. Images are then reconstructed from coincidence events. PET benefits from meter scale depth penetration, increased sensitivity due to a high level of emitted events, and millimeter length spatial resolution.

Low molecular weight (3 – 20 kDa) proteins and peptides are well-suited as PET imaging agents due to favorable pharmacokinetic properties, low immunogenicity, and rapid evolvability towards target binding through directed evolution. Small size enhances vascular extravasation through fenestrations between endothelial cells common to abnormally constructed tumor vessels, increasing tumor penetration and localization. Additionally, with molecular weights below the cutoff for glomerular filtration they are removed from the plasma rapidly, preventing deleterious background signal.[127] These

coupled effects provide an early time point imaging contrast not possible with larger molecules or traditional imaging modalities.[128–130]

Borrowing from the structural efficacy of the antibody and combining a stabilizing, core framework and secondary structure with diversifiable, affinity regions, several small protein 'alternative scaffolds' have been discovered and evolved as ligands towards numerous relevant biomarkers including vascular endothelial growth factor (VEGF), vascular endothelial growth factor receptor 2 (VEGFR-2), human epidermal growth factor receptor 2 (HER2), interleukin-6 (IL-6), and tyrosine-protein kinase Met (c-MET).[131,132]

### 1.5. Affibody and Gp2 scaffolds are promising biomarker imaging agents

Two compelling small proteins with potential as PET imaging agents for cancers that overexpress EGFR are the affibody and Gp2 scaffolds.

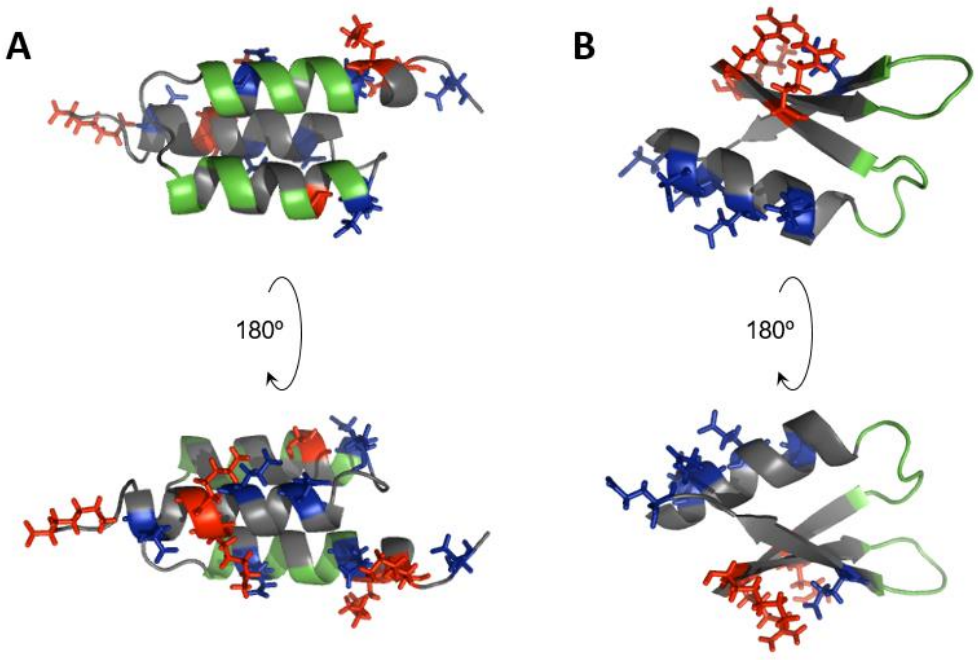



**Figure 1-1.** ***Solution structures of affibody and Gp2 scaffolds*** **(A) ABY-025 affibody and (B) wild-type Gp2 scaffolds with acidic (blue), basic (red), and paratope (green) residues highlighted [PDB: 1H0T[133] and 2WNM[134]]**

The affibody is a 58-residue (~ 6.5 kDa), naturally cysteine-free, three anti-parallel α-helical-bundle based on the Z domain, a thermally stabilized, reversible and rapidly refolding[135], and highly soluble variant of the B domain from *staphylococcal* protein A[136]. Thirteen surface-exposed residues on the first and second helices, naturally selected to bind the Fc region of immunoglobulin G, have been mutated through directed evolution towards a myriad of targets for imaging, therapeutic, and biotechnological purposes.[137] Affibody ligands for molecular targets of clinical concern have been explored extensively, resulting in low nanomolar to picomolar binders against TNF-α, EGFR, HER3, VEGF, VEGFR-2, IGF-1R, and IL-6.[138] As a rapidly evolving medical technology, many affibodies remain *in vitro* or preclinically evaluated, but clinical studies have been performed with $^{68}$Ga- and $^{111}$In-labeled anti-HER2 ABY-025 affibody[139–141] and molecular imaging studies of EGFR and IGF-1R with affibodies are ongoing[142,143].

Substantial effort to improve the affibody framework was undertaken by Feldwisch, et. al. in the context of an HER2-binding affibody, $Z_{HER2:342}$.[144] Primary considerations for an optimized framework included an increase in thermal stability and hydrophilicity, reduction of immunoglobulin affinity, and amenability for solid-phase peptide synthesis while retaining target binding. Site-directed mutagenesis of 11 framework positions provided a 4 ºC melting point ($T_m$) increase, reduced immunoglobin interaction, and enhanced hydrophilicity with negligible effects on HER2 affinity (ABY-025). Mutational benefits were often mutually exclusive. N23T raised the $T_m$ by 6 ºC, but this elevated stability was eradicated at the expense of increased hydrophilicity from A42S, A46S, and A54S. Mutations in ABY-025 increased the affibody framework charge state from +5/-7 to +6/-8, making acidic and basic residues 31% of the ligand backbone. Amino

acid based '$N_3S$' radiometal chelators have also been added to[145] or incorporated into[146] the affibody's N- and C-termini to modulate biodistribution with resultant reductions in gastrointestinal tract and excretory organs retention. Finally, in an effort to improve the affibody's tumor-to-background uptake ratio the third, nonbinding helix was removed completely with a concurrent addition of stabilizing cysteines for disulfide bridge formation.[147] Unfortunately, the resultant two-helical construct underperformed *in vivo* with a reduction in tumor uptake and increase in background signal suggesting the third-helix requisite for effective tumor targeting.[148] Given its highly charged framework, utility in medical and biotechnological applications, and initial structural stability, the affibody is a suitable protein with which to evaluate the impact of ionic charge mutations on physicochemical properties and *in vivo* ligand performance.

Gp2 is a 45-residue domain (~ 5.5 kDa) based on a dual-termini truncated version of the T7 phage gene 2 protein, an *E. coli* RNA polymerase inhibitor. It contains an α-helix opposing three β-strands with two flexible, solvent exposed loops typically diversified for epitope binding.[149] Gp2 ligands with low nanomolar affinity have been developed for model targets such as goat IgG, rabbit IgG, and lysozyme as well as clinically relevant markers including EGFR[149], insulin receptor[150], and PD-L1 (Kruziki, Sarma, and Hackel, under revision).

Framework mutations outside the paratope of Gp2 have resulted from error-prone PCR during directed evolution[149] or selected degeneracy applied at naïve library creation (Kruziki, Sarma, and Hackel, under revision). Positions 30, 31, and 32 shifted away from wild-type amino acids during rounds of PCR mutagenesis between target sorts, suggesting a region of mutational flexibility. Lead goat and rabbit IgG-binding clones retained several

of these residue replacements (E30V, Q32R and E30F, W31G) and demonstrated a concurrent $\Delta T_m$ of +3 or +13 ºC, respectively. Site-selective degeneracy in the Gp2 framework at three charged positions (E20, E27, and R44) revealed a preference for retaining wild-type charge residues over neutral amino acids.

Gp2 has segregated patches of acidic and basic residues on the α-helix and first and third β-strands respectively. These regions may drive nonspecific binding[73], increase off-target retention[38,79,80], and lower recoverable production yield due to decreased solubility[151–153]. Additionally, R42 and R44 are required for Gp2's natural function, *E. coli* RNA polymerase inhibition.[134] Therefore, arginine at these positions may be suboptimal for prokaryotic production and warrant replacement. As an alternative scaffold with demonstrable evolvability towards numerous cancer biomarkers and clinical potential as a noninvasive disease state imaging probe, an experimental evaluation of the effects surface charge reduction and redistribution have on the robustness and *in vivo* performance of the Gp2 scaffold are merited.

In the work described herein, we have used rational and semi-rational design in concert with combinatorial library selection to engineer the ionic charge of affibody and Gp2 scaffolds. The effect of charge on physicochemical properties and *in vivo* performance was evaluated in the context of anti-EGFR ligands bearing varied net, total, and distributive charge. Results aid future protein charge modulation and advance the therapeutic, diagnostic, and biotechnological benefit of alternative scaffolds.

# Chapter 2 - Synthetic and natural consensus design for engineering charge within an affibody targeting epidermal growth factor receptor

## 2.1. Outline

Protein ligand charge can impact physiological delivery with charge reduction often benefiting performance. Yet neutralizing mutations can be detrimental to protein function. Herein, three approaches are evaluated to introduce charged-to-neutral mutations of three cations and three anions within an affibody engineered to bind epidermal growth factor receptor. These approaches – combinatorial library sorting or consensus design, based on natural homologs or library-sorted mutants – are used to identify mutations with favorable affinity, stability, and recombinant yield. Consensus design, based on 942 affibody homologs, yielded a mutant of modest function ($K_d$ = 11 ± 4 nM, $T_m$ = 62 ºC, and yield = 4.0 ± 0.8 mg/L as compared to 5.3 ± 1.7 nM, 71 ºC, and 3.5 ± 0.3 mg/L for the parental affibody). Extension of consensus design to ten additional mutants exhibited varied performance including a substantially improved mutant ($K_d$ = 6.9 ± 1.4 nM, $T_m$ = 71 ºC, and 12.7 ± 0.9 mg/L yield). Sorting a homolog-based combinatorial library of 7 x $10^5$ mutants generated a distribution of mutants with lower stability and yield but did identify one strongly binding variant ($K_d$ = 1.2 ± 0.3 nM, $T_m$ = 69 ºC, and 6.0 ± 0.4 mg/L yield). Synthetic consensus design, based on the amino acid distribution in functional library mutants, yielded higher affinities ($p = 0.05$) with comparable stabilities and yields. The best of four analyzed clones had $K_d$ = 1.7 ± 0.5 nM, $T_m$ = 68 ºC, and 7.0 ± 0.5 mg/L yield. While all three approaches were effective in creating targeted affibodies with six charged-to-neutral mutations, synthetic consensus design proved to be the most robust. Synthetic

consensus design provides a valuable tool for ligand engineering, particularly in the context of charge manipulation.

## 2.2. Introduction

Ligand charge impacts in vivo performance, particularly physiological distribution. Cationic and highly anionic polypeptides clear the blood more rapidly than their neutral counterparts.[154] Additionally, positively charged proteins have increased vascular permeability and extravasate more rapidly in solid tumors.[75] Escalation in the number of positively or negatively charged residues increases kidney signal at early time points for multiple ligands.[80,82,155] In particular, reduction in the number of charged residues has been effective to improve biodistribution for several scaffolds including the affibody[80,155], knottin[79], and fibronectin type III domain[38]. Thus, charge modulation – in particular, reduction – warrants further investigation for improved ligand delivery.

Yet charge modification can significantly impact protein solubility, stability, affinity, specificity, and production. Introducing charged residues on the surface of a protein is generally well-tolerated and in some cases beneficial to stability and solubility.[90] However, removal of charged residues can be more problematic.[38,90] Numerous systems have exhibited reduced solubility upon charged-to-neutral residue replacements: two independent mutations (E48N or D130N) in type S1 dihydrofolate reductase[156], a combination of five neutralizing mutations in the KcsA potassium channel[151], K185F mutation in HIV integrase[152], K435L mutation in viral reverse transcriptase[157], and E100W mutation in leptin[153]. Some previous efforts to modify charge for pharmacokinetic impact studied large molecules, such as antibody Fc[71,158] in which a few amino acid mutations are

less likely to affect protein properties. Conversely, a similar number of charge mutations within a small ligand represents a greater fraction of the total protein and may therefore be more impactful.

The current study evaluates three different protein engineering approaches to modulate charge of the affibody domain. Affibodies are 58-residue, three-helix bundles based on the B domain of *staphylococcal* protein A.[136] Directed evolution of 13 residues on the surface of two helices has yielded molecular targeting agents for imaging and therapy, as well as scientific reagents.[159] Several affibodies have been engineered to bind selectively to epidermal growth factor receptor (EGFR) including $Z_{EGFR:1907}$, with 3 nM affinity[160]. EGFR is overexpressed in numerous cancers[161] and improved affinity ligands could have utility in molecular imaging and targeted therapy. The presented analyses were performed in the context of a hybrid affibody with the 13 paratope mutations of $Z_{EGFR:1907}$ and framework mutations engineered, in the context of $Z_{HER2:342}$[144], for improved synthesis and stability while maintaining binding affinity but without direct consideration of charge.

The use of natural homolog sequences to predict mutational tolerance has been useful[162–164], albeit inconsistent[165,166], for general mutations. We aimed to study the ability of homolog sequences to predict tolerance of charged-to-neutral mutations. Yet, consensus design is limited by the ability of nature to sample sequence space as well as the requisite retention of functional activity, which is different than our current activity of interest, *i.e.*, EGFR binding. Thus, considering the potential limitations of consensus design, we allowed homolog sequences to guide our selection of mutations while extending beyond natural consensus design. High throughput sorting of a combinatorial library – using the genotype-phenotype linkage of yeast display[167] and flow cytometry – expands the number of mutants

that can be efficiently evaluated albeit with potential limitations on the characterization of some phenotypes. The two approaches can be combined to overcome the limitations of natural consensus design. High throughput sorting can efficiently identify a population of functional sequences to serve as a synthetic basis for consensus design, which has been effectively implemented in evolving the chorismate mutase enzyme.[168]

Herein, these three approaches – consensus design, combinatorial library screening, and synthetic consensus design – were used to introduce numerous variations of six simultaneous charged-to-neutral mutations within an EGFR-binding affibody. Mutant analysis at the clonal and population levels provides significant insight into the impact of charged-to-neutral mutations within this affibody domain. While all three approaches yielded a range of affinities, stabilities, and recombinant yields including several highly functional mutants, synthetic consensus design was the most effective. A mutant with 3.2 ± 1.9-fold enhanced affinity (1.7 ± 0.5 nM $K_d$), comparable stability (68 ºC $T_m$), and improved yield (7.0 ± 0.5 mg/L), despite neutralization of six charged residues, was engineered.

## 2.3. Materials and Methods

### *2.3.1 Pfam Homolog Consensus*

Affibody EA68 (Figure 2-1, origin described in *Results*) was used as the seed sequence in a search of the Pfam database.[169] Proteins with less than 42 residues were discarded resulting in 942 homologs. Sequences were aligned to EA68 using a block substitution (BLOSUM) scoring matrix.[170] Amino acid frequencies were calculated at each site with a modification to account for repeat sequences without allowing their redundancy

to dominate sequence statistics: identical sequences were counted as the square root of the number of occurrences of the sequence.

### *2.3.2 Clone Production and Yield Quantification*

Genes were constructed by overlap extension PCR[171] of eight oligonucleotides (Integrated DNA Technologies) per clone, digested by NheI-HF and BamHI-HF restriction enzymes (New England Biolabs), and ligated into a pET-22b vector containing a C-terminal hepta-histidine (Novagen, EMD Millipore) with T4 DNA ligase (New England Biolabs). Plasmids were transformed via heat-shock into high efficiency T7 *E. coli* (New England Biolabs), and proper transformants were selected on lysogeny broth (10 g/L tryptone, 5.0 g/L yeast extract, 10.0 g/L sodium chloride in water) plates (15 g/L agar) containing kanamycin (50 mg/L). Two mL of saturated culture was added to 100 mL of lysogeny broth medium in a 250 mL shake flask, incubated at 37 °C, 250 rpm until an optical density at 600 nm between 0.5 and 1 units was achieved and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside for 2 hours at 25 – 28 °C. Cells were pelleted, resuspended in lysis buffer (50 mM sodium phosphate (pH 8.0), 0.5 M sodium chloride, 5% glycerol, 5 mM 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate, and 25 mM imidazole), and underwent four freeze-thaw cycles. The soluble fraction was isolated from cell lysate via centrifugation at 12,000*g* for 10 min. Affibodies were purified by immobilized metal affinity chromatography on 0.2 mL resin volume Ni-NTA HisPur spin columns (Thermo Fisher Scientific) and concentration assessed using absorbance at 280 nm on a Synergy H1 microplate reader (BioTek). Clones were further purified by reversed-phase high-performance liquid chromatography with a C18 column using a 15-

minute gradient of 25% to 80% elution buffer (90% acetonitrile, 9.9% water, 0.1% trifluoroacetic acid) and the remaining solution composed of running buffer (99.9% water, 0.1% trifluoroacetic acid). Isolated affibodies were lyophilized. Yields are presented as the average ± standard error of eight or nine production runs.

### *2.3.3 Library Construction and Sorting*

Genes for four charge-reduced affibody sub-libraries were constructed via overlap extension PCR of 12-15 oligonucleotides per sub-library, with degenerate codons present at six of the following positions: 2, 4, 7, 8, 15, 37, 49, 53, or 58 (Table 2-1).

**Table 2-1. Oligonucleotides used to create EA35 sub-libraries.**

| Sub-library 1 | |
|---|---|
| Oligonucleotide | Sequence |
| AmpF | CGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAG |
| Lib1F1 | TTCCAGACTACGCTCTGCAGGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAG |
| Lib1R1 | TGCGCTAGCCGACCCTCCGCCTCCGCTACCGCCT |
| Lib1F2A | GCGGAGGGTCGGCTAGCGCADYTGCCHHCTACGCCNYTDYTATGTGGGCCGCATGGGAA |
| Lib1F2B | GCGGAGGGTCGGCTAGCGCACAGGCCHHCTACGCCNYTDYTATGTGGGCCGCATGGGAA |
| Lib1F2C | GCGGAGGGTCGGCTAGCGCADYTGCCHHCTACGCCNYTCAGATGTGGGCCGCATGGGAA |
| Lib1F2D | GCGGAGGGTCGGCTAGCGCACAGGCCHHCTACGCCNYTCAGATGTGGGCCGCATGGGAA |
| Lib1R2A | GGTCATCTGCCAGCCGGTCAAGTTGGGCAAGTTACGAATCRMTTCCCATGCGGCCCACAT |
| Lib1R2B | GGTCATCTGCCAGCCGGTCAAGTTGGGCAAGTTACGAATCTGTTCCCATGCGGCCCACAT |
| Lib1F3 | TGACCGGCTGGCAGATGACCGCATTTATCGCCGCATTGGTTGACGACCCCAGCCAAAGCG |
| Lib1R3 | TGTGCGTCGTTCAATTTTTTTGCTTCTGCCAACAAGTTTGCGCTTTGGCTGGGGTCGTCA |
| Lib1F4 | AAAAAATTGAACGACGCACAAGCACCCHHCGGATCCGAACAAAAGCTTAT |
| AmpR | ATCTCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCC |

| Sub-library 2 | |
|---|---|
| Oligonucleotide | Sequence |
| AmpF | CGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAG |
| Lib1F1 | TTCCAGACTACGCTCTGCAGGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAG |
| Lib1R1 | TGCGCTAGCCGACCCTCCGCCTCCGCTACCGCCT |
| Lib2F2A | GCGGAGGGTCGGCTAGCGCADYTGCCHHCTACGCCAAADYTATGTGGGCCGCATGGGAA |
| Lib2F2B | GCGGAGGGTCGGCTAGCGCADYTGCCHHCTACGCCAAACAGATGTGGGCCGCATGGGAA |
| Lib2F2C | GCGGAGGGTCGGCTAGCGCACAGGCCHHCTACGCCAAADYTATGTGGGCCGCATGGGAA |
| Lib2F2D | GCGGAGGGTCGGCTAGCGCACAGGCCHHCTACGCCAAACAGATGTGGGCCGCATGGGAA |
| Lib1R2A | GGTCATCTGCCAGCCGGTCAAGTTGGGCAAGTTACGAATCRMTTCCCATGCGGCCCACAT |
| Lib1R2B | GGTCATCTGCCAGCCGGTCAAGTTGGGCAAGTTACGAATCTGTTCCCATGCGGCCCACAT |
| Lib1F3 | TGACCGGCTGGCAGATGACCGCATTTATCGCCGCATTGGTTGACGACCCCAGCCAAAGCG |
| Lib2R3 | TGCTTCTGCCAACAAGTTTGCGCTTTGGCTGGGGTCGTCAA |
| Lib2F4A | CAAACTTGTTGGCAGAAGCAKYGAAATTGAACGACGCACAAGCACCC |
| Lib2F4B | CAAACTTGTTGGCAGAAGCACAGAAATTGAACGACGCACAAGCACCC |
| AmpR | ATCTCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCC |

| Sub-library 3 | |
|---|---|
| Oligonucleotide | Sequence |
| AmpF | CGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAG |
| Lib1F1 | TTCCAGACTACGCTCTGCAGGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAG |
| Lib1R1 | TGCGCTAGCCGACCCTCCGCCTCCGCTACCGCCT |
| Lib1F3A | GCGGAGGGTCGGCTAGCGCADYTGCCHHCTACGCCNYTGAAATGTGGGCCGCATGGGAA |
| Lib1F3B | GCGGAGGGTCGGCTAGCGCACAGGCCHHCTACGCCNYTGAAATGTGGGCCGCATGGGAA |
| Lib1R2A | GGTCATCTGCCAGCCGGTCAAGTTGGGCAAGTTACGAATCRMTTCCCATGCGGCCCACAT |
| Lib1R2B | GGTCATCTGCCAGCCGGTCAAGTTGGGCAAGTTACGAATCTGTTCCCATGCGGCCCACAT |
| Lib3F3 | TGACCGGCTGGCAGATGACCGCATTTATCGCCGCATTGGTTGACGACCCCAGCCAAAGCG |
| Lib3R3 | TTTTTGCTTCTGCCAACAAGTTTGCGCTTTGGCTGGGGDDGTCAACCAATGCGGCGATAA |
| Lib3F4 | GTTGGCAGAAGCAAAAAAATTGAACGACGCACAAGCACCCHHCGGATCCGAACAAAAGCT |
| AmpR | ATCTCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCC |

| Sub-library 4 | |
|---|---|
| Oligonucleotide | Sequence |
| AmpF | CGACGATTGAAGGTAGATACCCATACGACGTTCCAGACTACGCTCTGCAG |
| Lib1F1 | TTCCAGACTACGCTCTGCAGGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAG |
| Lib1R1 | TGCGCTAGCCGACCCTCCGCCTCCGCTACCGCCT |
| Lib1F3A | GCGGAGGGTCGGCTAGCGCADYTGCCHHCTACGCCNYTGAAATGTGGGCCGCATGGGAA |
| Lib1F3B | GCGGAGGGTCGGCTAGCGCACAGGCCHHCTACGCCNYTGAAATGTGGGCCGCATGGGAA |
| Lib1R2A | GGTCATCTGCCAGCCGGTCAAGTTGGGCAAGTTACGAATCRMTTCCCATGCGGCCCACAT |
| Lib1R2B | GGTCATCTGCCAGCCGGTCAAGTTGGGCAAGTTACGAATCTGTTCCCATGCGGCCCACAT |
| Lib1F3 | TGACCGGCTGGCAGATGACCGCATTTATCGCCGCATTGGTTGACGACCCCAGCCAAAGCG |
| Lib4R3 | GTTCAATTTTTTTGCTTCTGCCAACAAGTTTGCGCTTTGGCTGGGGTCGTCA |
| Lib4F4 | CAGAAGCAAAAAAATTGAACHHCGCACAAGCACCCHHCGGATCCGAACAAAAGCTTAT |
| AmpR | ATCTCGAGCTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGGATCC |

PCR reaction products and linearized pCT surface display vector were homologously recombined upon transformation via electroporation into EBY100 yeast.[172] Transformants were allowed to multiply in SD-CAA (0.07 M sodium citrate (pH 5.3), 6.7

g/L yeast nitrogen base, 5 g/L casamino acids, and 20 g/L glucose in water) at 30 °C, 250 rpm until log phase growth, then pelleted and resuspended in SG-CAA (0.1M sodium phosphate (pH 6.0), 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 19 g/L galactose, and 1 g/L glucose in water) and shaken at 30 °C, 250 rpm for 8-24 h to induce cell surface affibody expression. Libraries were then pooled for fluorescence-activated cell sorting experiments. Two populations of 6.7 x $10^6$ (10x theoretical diversity) cells were pelleted, washed with 4 °C PBS containing 1% bovine serum albumin (PBSA), and heated to 70 °C for 30 min. Cells were kept at 4 °C from this point forward. The yeast were then washed with cold PBSA, pelleted, and resuspended in PBSA with 20 mg/L mouse anti-c-MYC antibody (clone 9E10, Covance) with or without 20 nM biotinylated EGFR ectodomain (Prospec). After incubating on ice for 10 min, the populations were washed, pelleted, and resuspended in PBSA, streptavidin-Alexa Fluor 647 (Life Technologies), and goat anti-mouse-FITC antibody (Sigma-Aldrich). Yeast were sorted on a FACSAria II (BD Biosciences) flow cytometer with the top 5% of cells, in terms of EGFR binding : surface expression ratio, collected. Plasmids from collected yeast were recovered with a Zymoprep kit (Zymo Research) following manufacturer's protocol and affibody-encoding genes amplified by PCR. Resultant PCR products were homologously recombined with pCT vector upon electroporation into EBY100 yeast. Two additional sorts were performed following this procedure, with the exception that in those cases cells were heat-treated at 80 °C rather than 70 °C. Encoding regions from recovered yeast from the final sort were PCR amplified, ligated into pET-22b vector, transformed into high efficiency 5-alpha *E. coli* (New England Biolabs) and grown on plates (lysogeny broth with 15 g/L agar and 50 mg/L kanamycin). Genes from fifty-four plate colonies were sequenced and found to all

encode for affibodies. Sixteen sequences had combinations of charge mutations not designed into the sub-libraries – in some cases more neutralizations and in other cases less – and were discarded. The remaining thirty-eight sequences were used to create synthetic consensus frequencies.

### *2.3.4 Circular Dichroism*

Purified, lyophilized affibody clones were resuspended in PBS to a concentration of 0.5 - 1 g/L. Ellipticity was measured between 200 and 260 nm wavelengths at room temperature on a Jasco J-815 spectrophotometer in a quartz cuvette with 1 mm path length. Thermal denaturation was performed by measuring ellipticity at 220 nm from 20 to 98 °C (2 °C/min). The midpoint of thermal denaturation ($T_m$) was calculated using a two-state unfolding model. Heated samples were cooled to 20 ºC and subjected to a post-heating wavelength scan.

### *2.3.5 Affinity Measurement*

A431 epidermoid carcinoma cells, which express approximately 2.8 x $10^6$ EGFR per cell[173], were provided by Dr. Daniel Vallera (University of Minnesota). Cells were cultured in high-glucose Dulbecco's modified Eagle's medium with 10% fetal bovine serum and 1% penicillin streptomycin at 37 °C in humidified air with 5% $CO_2$. At 70-80% confluence, cells were detached with trypsin-ethylenediaminetetraacetic acid, pelleted, and resuspended in cold PBSA to prevent EGFR internalization prior to use. Fifty thousand cells were washed with PBSA and incubated with varying concentrations of ligand. Cells were pelleted, washed again with PBSA, and incubated with fluorescein-conjugated anti-$His_6$ antibody (Abcam) in PBSA. Cells were washed and analyzed using flow cytometry

on an Accuri C6 (BD Biosciences). Median fluorescence values at each ligand concentration, corresponding to binding fraction, were used to determine affinity values with least squares regression assuming a 1:1 binding interaction. Equilibrium dissociation constants, $K_d$, are presented as the average ± standard error of three to five titrations.

### *2.3.6 Size Exclusion Chromatography and Target Specificity Assay*

Purified and lyophilized affibody clones, aprotinin, and cytochrome C were resuspended in PBS to a concentration of 1 mg/mL. 200 uL of each sample was run through a Superdex75 10/300 GL size exclusion column on an ÄKTAprime Plus chromatography system (GE Healthcare Life Sciences) at a flow rate of 0.5 mL/min for 1 hour. Retention volumes were determined by monitoring light absorbance at 280 nm.

MCF-7 breast adenocarcinoma cells which express fewer than 5 x $10^3$ EGFR per cell, graciously provided by Dr. Deepali Sachdev (University of Minnesota), were used as pseudo-negative controls for target specificity assays. MCF-7 cells were cultured, detached, and processed identically to A431. After detachment and PBSA washing, fifty thousand A431 or MCF-7 cells were incubated with 500 nM affibody ligand. Upon binding equilibrium, cells were pelleted, washed with PBSA and incubated with fluorescein-conjugated anti-$His_6$ antibody is PBSA. Cells were pelleted, washed in PBSA again then analyzed via flow cytometry on an Accuri C6. Median fluorescence with background signal subtracted are presented as the mean ± standard error of three replicates.

## 2.4. Results

### *2.4.1 Hybridizing EGFR Targeting Mutations with an Improved Framework*

Affibody $Z_{EGFR:1907}$ engineered by Friedman et al.[160] to target EGFR was the partial basis for our starting ligand. Seven framework mutations, V1A, D2E, N3A, F5Y, N6A, N23T, and S33A, identified by Feldwisch et al.[144] in the context of a HER2 targeting affibody ZHER2:342, were implemented to improve stability and reduce immunoglobulin binding (Figure 2-1).

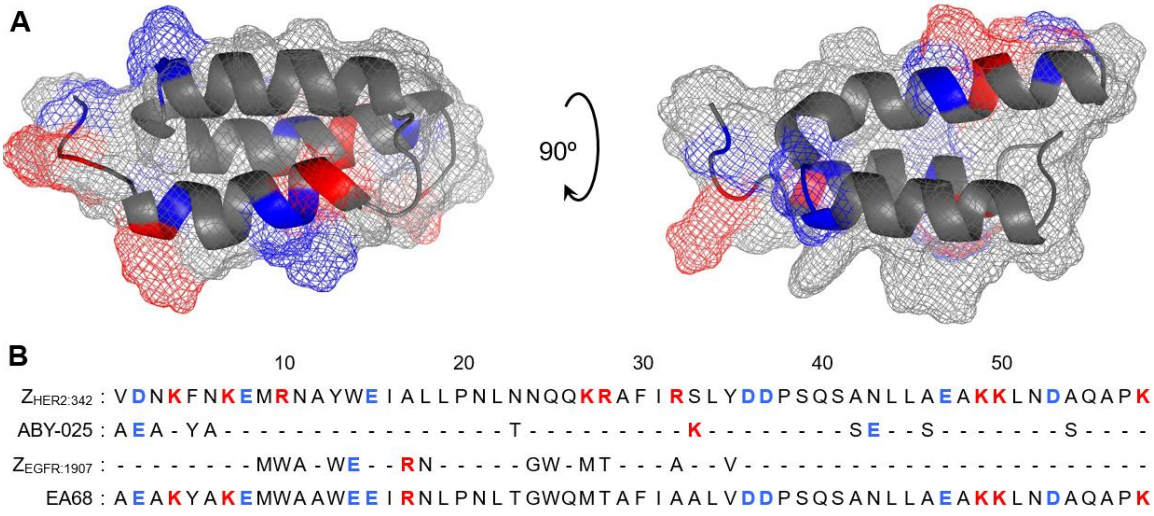



**Figure 2-1.** ***Affibody domain solution structure and affibody variant sequences used to create clone EA68*** **(A) Schematic of affibody domain variant (PDB: 1H0T[133]). (B) ZHER2:342[174] is an evolved HER2 binder with the original framework sequence. ABY-025[144] is a mutant with an optimized framework. $Z_{EGFR:1907}$[160] is a mutant with 13 sites evolved to bind EGFR. EA68 is a combined mutant with the EGFR binding mutations and the optimized framework, with five exceptions: three A to S mutants – at sites 42, 46, and 54 – were retained as A. These mutations were selected for protein synthesis but slightly destabilize the protein; the S33K mutation was avoided because of the introduction of charged K and S33A was chosen, which provides 10 ºC with no impact on HER2 affinity and reduced Ig binding; the N43E mutation was avoided because of the introduction of charged E. Negatively and positively charged residues are colored in blue and red, respectively, in the sequence. The corresponding sites in the structure are also colored but note that the structure does not correspond to this amino acid sequence, but rather a different variant.**

There are five exceptions from the optimized framework as published: three A to S mutants – at sites 42, 46, and 54 – were retained as alanine. These mutations were selected to improve protein synthesis at the expense of a slight reduction in stability; the S33K and N43E mutations were avoided because of the introduction of charged groups; S33A was chosen to provide stabilization ($T_m$ increases 10 ºC) with no impact on HER2 affinity and reduced Ig binding. This hybrid molecule, termed EA68 for *E*GFR-targeting *a*ffibody with *six* positively charged residues and *eight* negatively charged residues, was the basis for our protein engineering efforts. EA68 exhibits binding affinity to EGFR-overexpressing A431 cells with $K_d$ = 5.3 ±1.7 nM (Figure 2-2A), which is comparable to $Z_{EGFR:1907}$ (2.8 nM). The molecule exhibits the expected α-helical secondary structure (Figure 2-2B). The midpoint of thermal denaturation is 71 ºC (Figure 2-2C), which is comparable to the HER2 binding affibody with a similarly optimized framework ($T_m$ = 69 ºC) but different sequence in the binding paratope. Unfolding is only partially reversible as α-helical secondary structure, but varied signal, is observed upon cooling to 25 ºC after thermal denaturation (Figure 2-2B). The recombinant yield from *E. coli* is 3.5 ± 0.3 mg/L in unoptimized shake flask productions. Thus, EA68 serves as a useful baseline to initiate charge engineering.

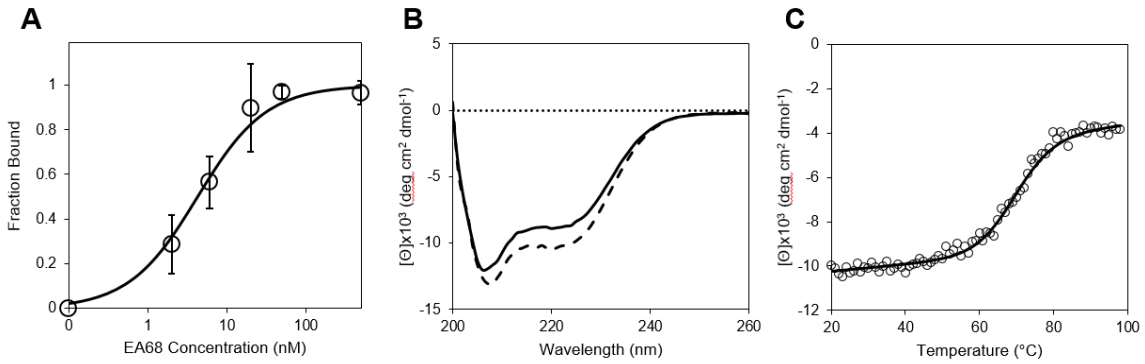



**Figure 2-2.** ***Characterization of EA68 affibody.*** **(A) A431 cells overexpressing EGFR were incubated on ice with the indicated concentration of purified EA68 followed by fluorescein-conjugated anti-$His_6$ antibody and analyzed by flow cytometry. The equilibrium dissociation constant ($K_d$) is calculated by minimizing sum of squared errors between experiment and a theoretical 1:1 binding model. (B,C) Purified EA68 was analyzed by circular dichroism spectroscopy by (B) scanning wavelengths from 200 to 260 nm at 20 ºC. Triplicate scans were collected before and after thermal denaturation and cooling and (C) scanning temperatures from 20 ºC to 98 ºC (2 ºC/min.) at 220 nm. The midpoint of thermal denaturation was calculated from a two-state model.**

### *2.4.2 Charge Removal by Consensus Design*

A consensus design approach was implemented to identify tolerable mutations to remove charged residues. Natural homologs to the engineered affibody were identified from the Pfam[169] database using EA68 as the basis sequence. Amino acid frequencies were calculated at each site based on the 942 homologs in the family. Identical sequences were partially discounted to account for their multiple natural appearances without allowing their redundancy to dominate sequence statistics. The charged framework sites exhibit variable conservation with a frequency range of 5 - 92%, median of 64%, and standard deviation of 30% (Table 2-2). Frequencies of neutral residues range from 2% to 72%. The median number of different amino acids observed at each charged site was 6.5. The median Shannon entropy[175] was 1.2 for all sites and 1.4 at charged framework sites. These values

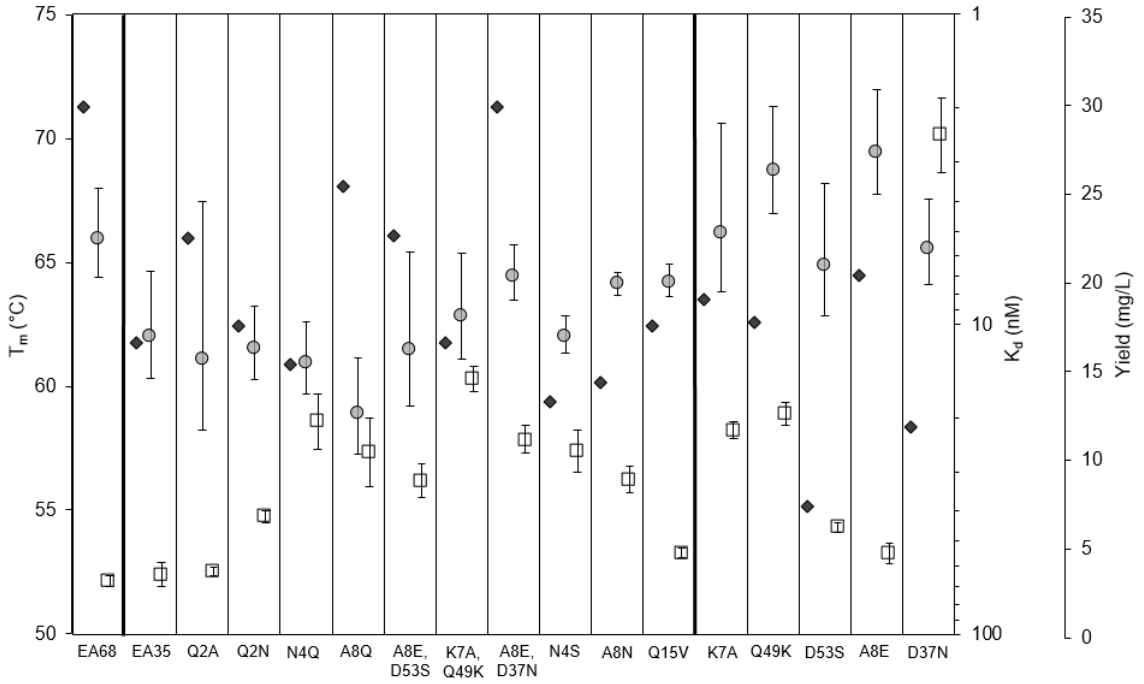
indicate a modest level of diversity from which to design mutants. Three positive and three negative residues were mutated to neutral within the framework of EA68. Consensus design identified the set of six charged-to-neutral mutations whose summed natural frequencies were highest: E2Q (24%), K4N (38%), E8A (21%), E15Q (30%), K49Q (21%), and K58N (62%). A clone incorporating all six mutations, named EA35, exhibited a comparable affinity (11 ± 4 nM) and recombinant yield (4.0 ± 0.8 mg/L) to EA68 but a 10 ºC reduction in thermal stability (Figure 2-3, Figure 2-13, and Figure 2-14).

**Table 2-2. Natural amino acid frequencies at each charged site (excluding sites 14 and 17 engineered for binding). Amino acids for clone EA68 and the mutant that removes three acidic and three basic residues while maximizing natural neutral frequency (Max Neutral, also known as EA35) are reported. The amino acids of mutants are noted with matches to EA35 indicated as •.**

| | 2 | 4 | 7 | 8 | 15 | 36 | 37 | 47 | 49 | 50 | 53 | 58 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 17% | | 10% | 21% | | | | | | 5% | 2% | |
| C | | 1% | | | | | | | | | | |
| D | 36% | 6% | 2% | 23% | 1% | 80% | 78% | 1% | | | 80% | |
| E | 5% | 1% | 19% | 38% | 55% | 18% | | 92% | 1% | | 2% | 5% |
| F | 1% | | | | | | | | | | | |
| G | 1% | 1% | 1% | | | | | | | | | |
| H | 1% | | | | | | 11% | | 2% | 1% | | |
| I | | | | | | | | | 2% | | | 3% |
| K | 3% | 24% | 45% | 2% | | | | | 72% | 87% | | 18% |
| L | | | | | 1% | | | | | | | |
| M | | | | | | | | | | | | |
| N | 12% | 38% | 4% | 5% | 1% | 2% | 11% | | | 2% | 1% | 62% |
| P | | | 2% | | | | | | | | | |
| Q | 24% | 16% | 5% | 11% | 30% | | | 5% | 21% | 1% | 1% | |
| R | | 10% | 1% | | 2% | | | | 1% | 3% | | 5% |
| S | | 2% | | | 1% | | | 1% | | 1% | 14% | 8% |
| T | | 1% | 9% | | | | | 2% | | | | |
| V | 1% | | 1% | | 10% | 1% | | | | | | |
| W | | | | | | | | | | | | |
| Y | | | | | | | | | | | 1% | |
| EA68 | E | K | K | E | E | D | D | E | K | K | D | K |
| Max Neutral | Q | N | K | A | Q | D | D | E | Q | K | D | N |
| EA35 | • | • | • | • | • | • | • | • | • | • | • | • |
| EA35 Q2A | ***A*** | • | • | • | • | • | • | • | • | • | • | • |
| EA35 Q2N | ***N*** | • | • | • | • | • | • | • | • | • | • | • |
| EA35 N4Q | • | ***Q*** | • | • | • | • | • | • | • | • | • | • |
| EA35 A8Q | • | • | • | ***Q*** | • | • | • | • | • | • | • | • |
| EA35 A8E/D53S | • | • | • | ***E*** | • | • | • | • | • | • | ***S*** | • |
| EA35 K7A/Q49K | • | • | ***A*** | • | • | • | • | • | ***K*** | • | • | • |
| EA35 A8E/D37N | • | • | • | ***E*** | • | • | ***N*** | • | • | • | • | • |
| EA35 N4S | • | ***S*** | • | • | • | • | • | • | • | • | • | • |
| EA35 A8N | • | • | • | ***N*** | • | • | • | • | • | • | • | • |
| EA35 Q15V | • | • | • | • | ***V*** | • | • | • | • | • | • | • |
| EA35 K7A | • | • | ***A*** | • | • | • | • | • | • | • | • | • |
| EA35 Q49K | • | • | • | • | • | • | • | • | ***K*** | • | • | • |
| EA35 D53S | • | • | • | • | • | • | • | • | • | • | ***S*** | • |
| EA35 A8E | • | • | • | ***E*** | • | • | • | • | • | • | • | • |
| EA35 D37N | • | • | • | • | • | • | ***N*** | • | • | • | • | • |

An additional ten mutants, similarly selected based on summed natural amino acid frequencies, were produced and evaluated (Figure 2-3). Overall, all but one mutant (EA35 A8Q) had affinities within 3-fold of EA68 (range: 6.9 ± 1.4 nM to 20 ± 7 nM versus 5.3 ± 1.7 nM). All recombinant yields were comparable or better than EA68 (range: 4.0 ± 0.8 mg/L to 16.6 ± 0.8 mg/L versus 3.5 ± 0.3 mg/L). However, 9 of 11 thermal stabilities decreased by at least 5 ºC (range: 59 ºC to 71 ºC versus 71 ºC). The reversibility of unfolding varies between mutants in regards to secondary structure as assessed by circular

dichroism spectroscopy (Figure 2-4). The optimal combination of affinity and stability is EA35 A8E/D37N: 6.9 ± 1.4 nM affinity and 71 ºC $T_m$, which had the fifth highest sum of natural frequencies.




**Thermal stability (black diamonds), affinity (gray circles), and recombinant yield (white squares) for EA68, EA35, and EA35 variants. The midpoints of thermal denaturation ($T_m$) were measured by circular dichroism spectroscopy of purified protein (n = 1 - 2 with a mean difference of 7%). The equilibrium dissociation constants were measured by titration of affibody mutants on A431 cells using flow cytometry to detect binding (n = 3 - 5). Recombinant yield is the amount of purified affibody recovered per L of culture (n = 8 - 9). Points and error bars represent means and standard errors.**

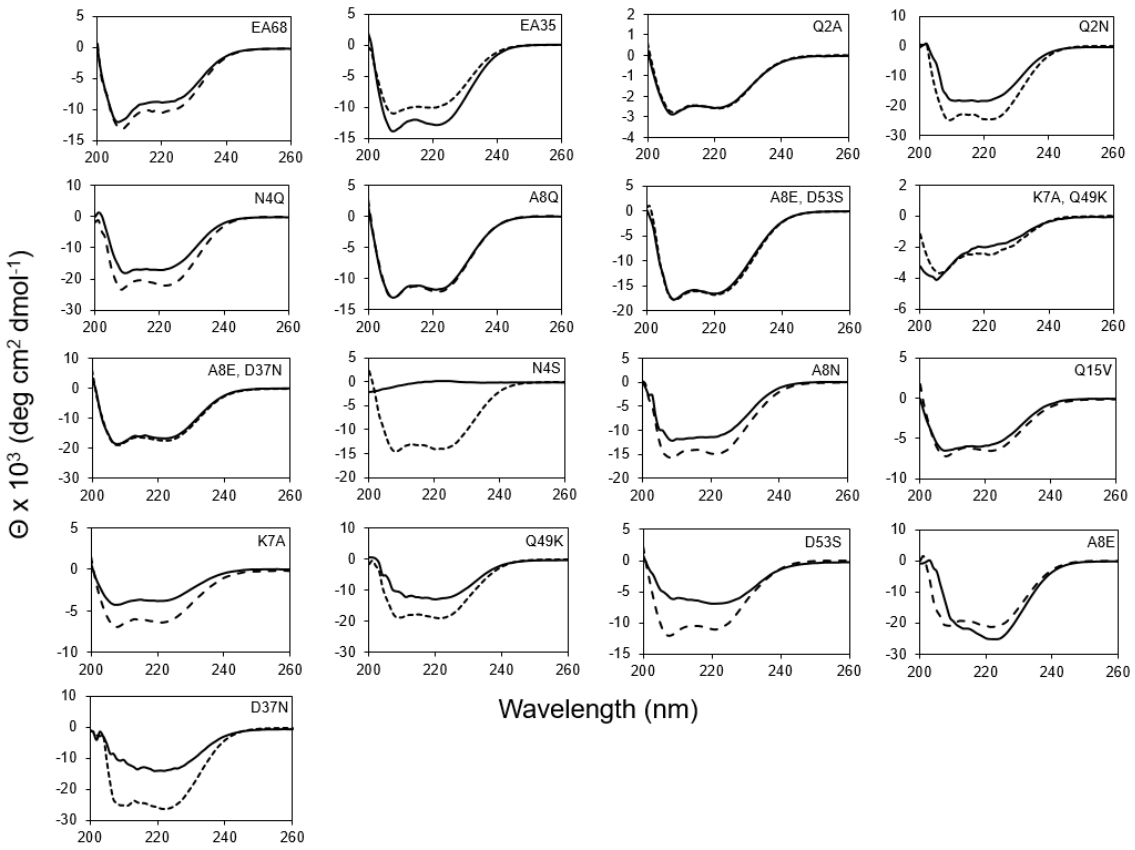



**Figure 2-4. *Circular dichroism spectroscopy of mutants.* Purified affibody domains were analyzed by circular dichroism spectroscopy by scanning wavelengths from 200 to 260 nm. Triplicate scans were collected before (- - -) and after (——) thermal denaturation and cooling.**

Several sets of mutants differ by the choice of neutral amino acid at a single site, which enables sitewise analysis of amino acid choices at sites 2, 4, 8, and 15 (Figure 2-3 and Figure 2-5). At site 2, A, N, and Q yield indistinguishable affinities while Q2A provides modestly elevated stability. At site 4, N, Q, and S yield similar affinities and stabilities; N4Q and N4S provide elevated yields. At site 8, A8N provides the optimal affinity whereas A8Q provides the highest stability; A8Q and A8N provide elevated yields. Q and V, at site 15, provide similar affinities, stabilities, and yields. Moreover, three mutants have a charge reversion relative to EA35 as well as a neutralizing mutation (i.e.

A8E/D53S, K7A/Q49K, and A8E/D37N) which required the production and analysis of five additional mutants to allow for individual site comparisons (Figure 2-3 and Figure 2-5). EA35 K7A/Q49K maintains affinity and stability with high yield. Analysis of the individual mutations to EA35 revealed that both EA35 K7A and EA35 Q49K had improved affinity, moderate stability, and high yield. EA35 A8E/D53S has modest affinity but improved stability and yield. Single mutant EA35 A8E had high affinity and stability and reasonable recombinant yield while EA385 D53S bound with 6.4 ± 2.9 nM affinity but was less stable. While EA35 A8E/D37N was the most stable EA35 mutant ($T_m$ = 71 ºC), EA35 A8E was only slightly more stable than EA35 (65 ºC) and EA35 D37N was less stable (58 ºC). The synergy of these mutations is reasonable given their proximity (11.2 Å, on average, in affibody variants with solved structures.). Yet binding and yield lack these synergies as EA35 A8E provides the strongest binding (2.8 ± 1.0 nM) while EA35 D37N has the highest yield (32.3 ± 2.4 mg/L).

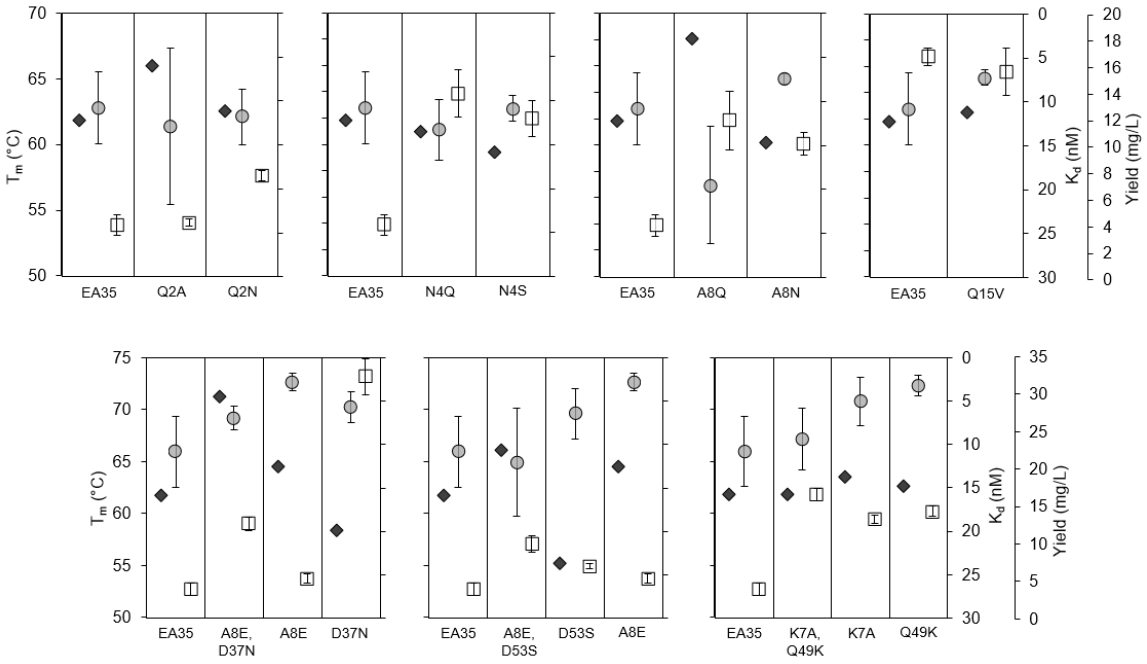



**Figure 2-5.** ***Sitewise comparisons from Figure 2-3.*** **The midpoints of thermal denaturation ($T_m$, diamonds) were measured by circular dichroism spectroscopy of purified protein (n = 1 - 2 with reproducibility of 7%). The equilibrium dissociation constants (circles) were measured by titration of affibody mutants on A431 cells using flow cytometry to detect binding (n = 3 - 5). Recombinant yield (squares) is the amount of purified affibody recovered per L of culture (n = 8 - 9). Points and error bars represent means and standard deviations.**

The properties of the collection of consensus mutants were analyzed to assess the utility of natural sequence frequency to guide mutational design. No significant correlation was observed between performance (affinity, stability, or yield) and natural sequence frequency as assessed by the change in natural frequency between pairwise mutants (Figure 2-6A-C) or the cumulative frequencies of neutral mutant amino acids (Figure 2-6D-F).

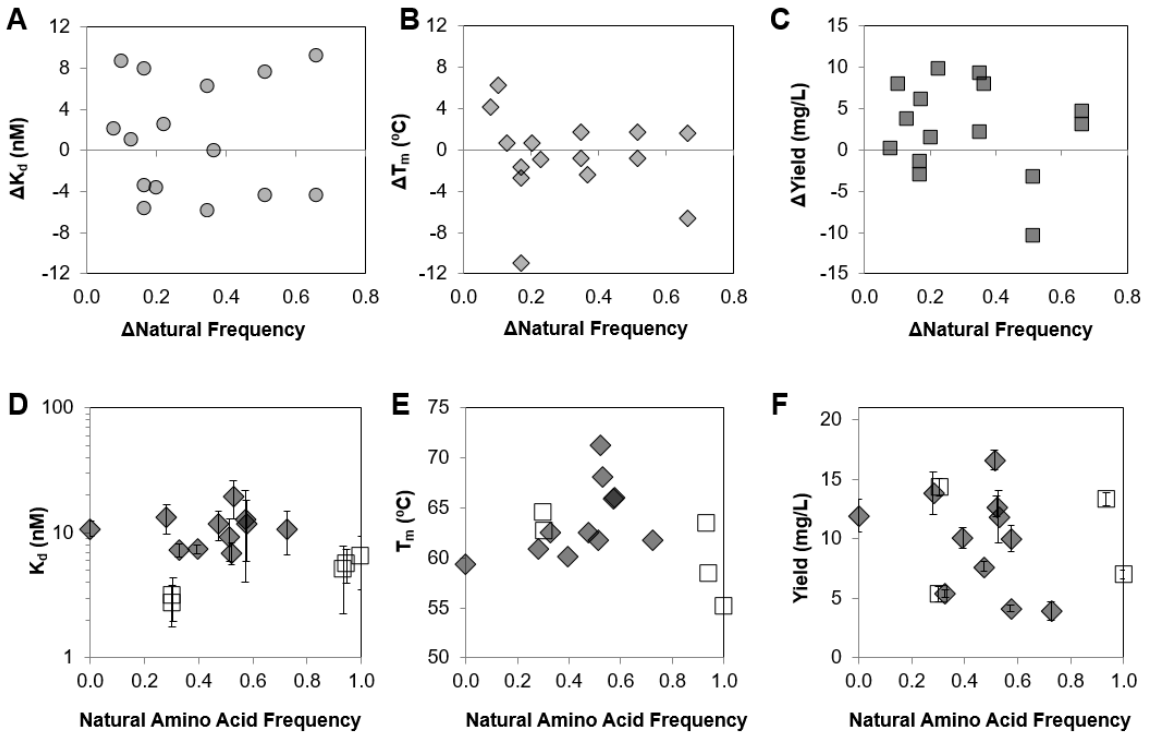



**Figure 2-6.** ***Relationships between performance (affinity, stability, and yield) and amino acid frequency in natural homologs.*** **(A-C) For sets of mutants differing at a single site, the affinity (A), thermal stability (B), and recombinant yield (C) are plotted against the change in natural frequency of the amino acids at the site of interest. (D-F) For EA35 and its equally charged mutants, the affinity (D), thermal stability (E), and recombinant yield (F) are plotted against the normalized sum of natural frequencies at the six sites mutated relative to EA68. p > 0.05 for all plotted relationships.**

### *2.4.3 Charge Removal in a Combinatorial Library*

Based on the variability of consensus design for EA68 and other molecules[165,166], a parallel approach was used that expanded the evaluated sequence space. Combinatorial libraries were created in which three positively charged sites and three negatively charged sites were mutated to neutral amino acids (Table 2-3). The six mutated sites were chosen again based on natural homolog sequence frequencies, but now using cumulative neutral residue frequency as the metric. Thus, the first library design diversified E2, E8, and E15 (56%, 37%, and 42% neutral naturally) while conserving D36, D37, E47, and D53 (2%,

22%, 7%, 18% neutral naturally); analogously, K4, K7, and K58 (59%, 33%, and 72% neutral) were diversified while K49 and K50 (25% and 10%) were conserved. Three variations were also considered in which an alternative site was diversified and, concomitantly to retain the same overall charge, a previously diversified site was conserved as charged: (Library 1) K49 was diversified and K7 conserved; (Library 2) D37 was diversified and E8 conserved; or (Library 3) D53 diversified and E8 conserved. The permitted neutral residues were chosen as degenerate codons that incorporated the naturally frequent amino acids while also permitting additional chemical functionalities yet excluding charged residues. Five to nine amino acids were permitted at each site to yield theoretical diversities of $\leq 2 \times 10^5$ mutants. The four libraries were constructed within a yeast surface display system with $2.1 \times 10^7$ to $3.4 \times 10^7$ transformants per library, yielding > 99% sampling of theoretical sequence space.[176]

**Table 2-3. Amino acid diversities allowed in the four sublibraries for selection of EA68 mutants with removal of three acidic and three basic residues.**

| Sub-library | E2 | K4 | K7 | E8 | E15 | D36 | D37 | E47 | K49 | K50 | D53 | K58 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ASTIFVQ | FLSYPHITN | ASTFLPIV | ASTIFVQ | VALSQ | D | D | E | K | K | D | FLSYPHITN |
| 2 | ASTIFVQ | FLSYPHITN | K | ASTIFVQ | VALSQ | D | D | E | VALSQ | K | D | FLSYPHITN |
| 3 | ASTIFVQ | FLSYPHITN | ASTFLPIV | E | VALSQ | D | FLSYPHITN | E | K | K | D | FLSYPHITN |
| 4 | ASTIFVQ | FLSYPHITN | ASTFLPIV | E | VALSQ | D | D | E | K | K | FLSYPHITN | FLSYPHITN |

The combinatorial libraries were pooled and sorted for stability and binding affinity by thermally treating the yeast-displayed proteins at 70 - 80 ºC and, after cooling, testing binding to 20 nM EGFR ectodomain. Note that thermal denaturation of EA35 and its variants is not fully reversible in the yeast displayed context, which enables this thermal stress to be effectively used (Figure 2-7A). Mutants exhibiting the best retention of thermal stability and binding were collected and sequenced. Sitewise amino acid frequencies (Figure 2-7B,C) exhibit a range of behaviors. Amino acids dominant in stable binders are

frequent in natural homologs although the inverse is not always true. For example, site 58 is 86% N when naturally neutral, but only 5% N in synthetically sorted mutants. No synthetically dominant residue is rare naturally, although several residues absent naturally are observed at reasonable percentages in sorted binders (maximum: 27% and median: 4% for amino acids not observed in natural homologs).

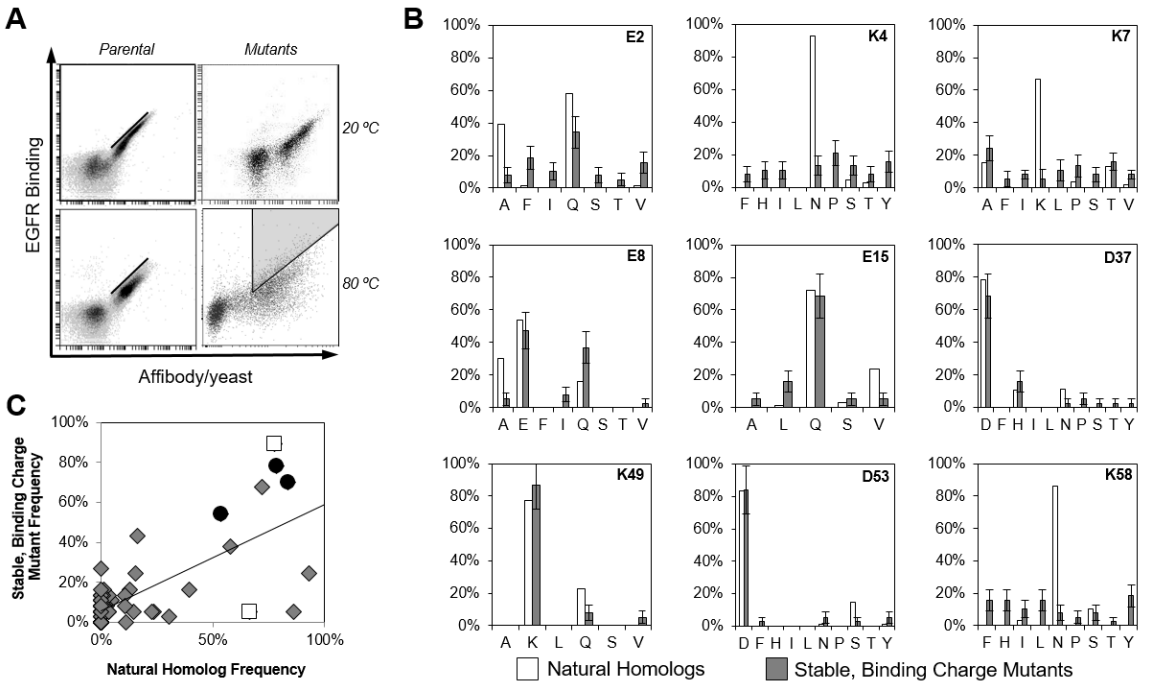



**Figure 2-7.** ***Combinatorial library analysis.*** **(A) Yeast displaying affibodies (left: parental EA68, right: enriched stable binding mutants) were treated at 20 ºC or 80 ºC for 30 min, cooled, and incubated with 20 nM biotinylated EGFR ectodomain. Binding was detected by streptavidin-AlexaFluor 647, and affibody display was quantified with anti-c-myc antibody with anti-mouse-AlexaFluor 488 using flow cytometry for quantification. (B) Amino acid frequencies at each diversified site in the combinatorial libraries and enriched stable binders. (C) The observed sequence frequencies in the stable binders relative to the amino acid frequencies in natural affibody homologs. Squares represent K7 and K49 for which the synthetic sequence frequencies sum to, at most, 100% because of the library design requirement to maintain three cations. Circles indicate E8, D37, and D53 for which the synthetic sequence frequencies sum to, at most, 200%.**

Four mutants from the sorted population were produced recombinantly and evaluated for yield, affinity, stability, and secondary structure (Figure 2-8). Three of the four clones have thermal stabilities lower than any of the consensus designed mutants (Figure 2-9). The average $T_m$ of mutants selected from the library is 57 ± 4 ºC versus 64 ± 1 ºC for consensus designed mutants (p = 0.23 for two-tailed Student's t-test). Production yields were slightly lower for library mutants (6.7 ± 2.0 mg/L) than consensus mutants (9.8 ± 1.2 mg/L; p = 0.24). Affinities were similar between groups (12 ± 6 nM for library mutants and 11 ± 1 nM for consensus mutants). Notably, though, one of the mutants (EA35 Q2A/K7T/A8E/Q49K/D53Y/N58Y) exhibited very good stability (69 ºC) and affinity (1.2 ± 0.3 nM) with an acceptable yield (6.0 ± 0.4 mg/L).

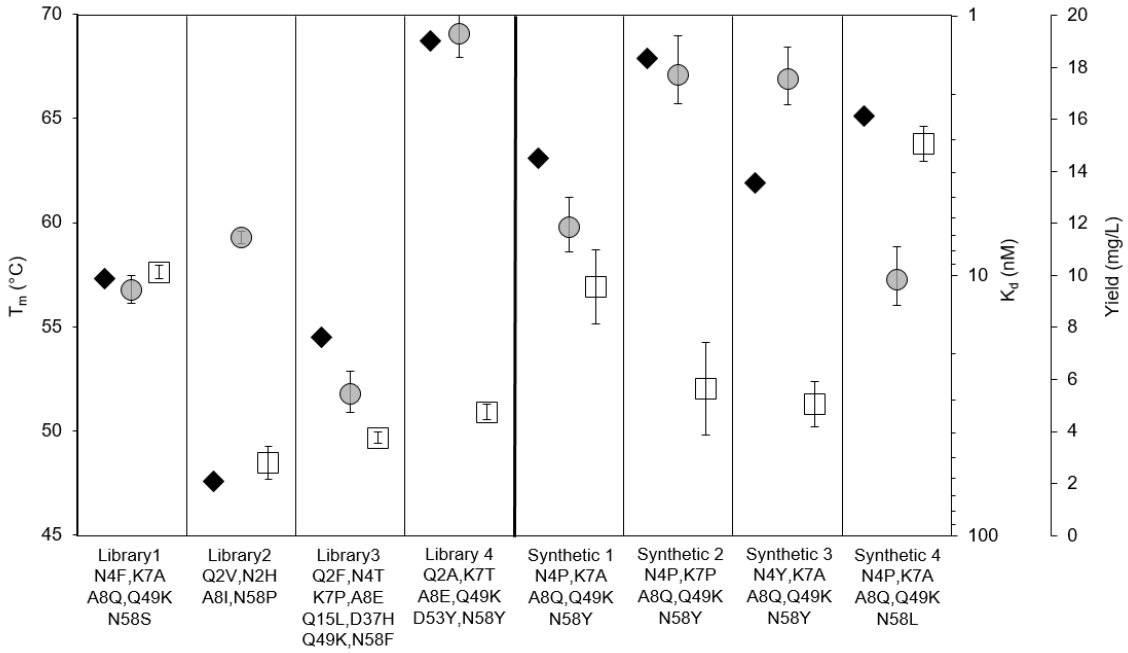



**Figure 2-8.** ***Analysis of clones from combinatorial library and synthetic consensus design.*** **The midpoints of thermal denaturation ($T_m$) were measured by circular dichroism spectroscopy of purified protein. The equilibrium dissociation constants were measured by titration of affibody mutants on A431 cells using flow cytometry to**

**detect binding (n = 3). Recombinant yield is the amount of purified affibody recovered per L of culture (n = 9). Points and error bars represent means and standard error.**

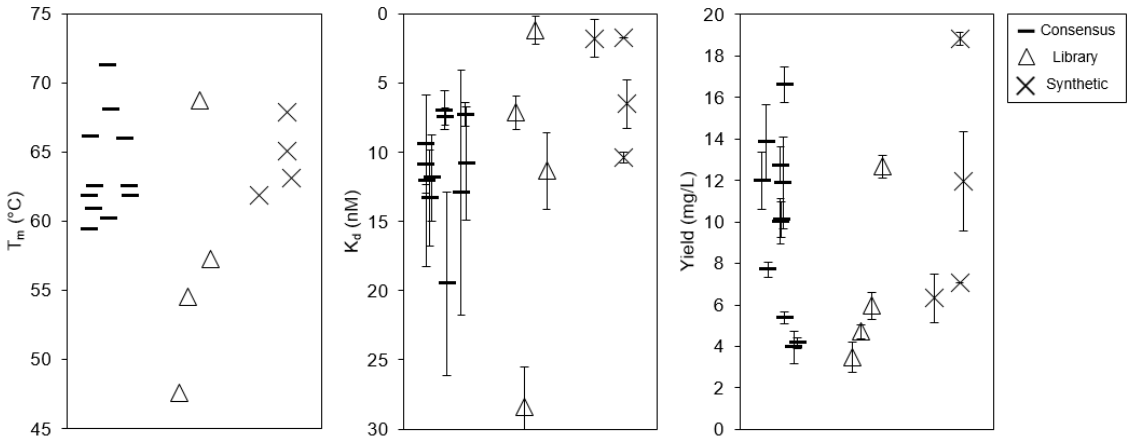



**Figure 2-9. *Relationship between biophysical properties and design strategy.* Thermal stability (left), affinity (center), and recombinant yield (right) comparison across design strategies.**

### *2.4.4 Consensus Design from a Synthetic Population*

The amino acid frequencies observed in the stable, binding population were used as the basis for synthetic consensus design. Such an approach – using a synthetic source of diversity for consensus design – was successful for enzyme engineering.[168] As in the consensus design from natural homolog sequences above, mutants were designed to neutralize three negatively charged residues and three positively charged residues with a maximal sum of observed frequencies. The top four synthetic consensus designs were produced and evaluated (Figure 2-8). Thermal stabilities were comparable to consensus design (65 ± 1 ºC versus 64 ± 1 ºC); yields were also similar (11 ± 3 mg/L versus 9.8 ± 1.2 mg/L). Affinities were significantly improved ($p = 0.05$) with two improved mutants (EA35 N4Y/K7A/A8Q/Q49K/N58Y with $K_d$ = 1.8 ± 0.4 nM and EA35 N4P/K7P/A8Q/Q49K/N58Y with $K_d$ = 1.7 ± 0.5 nM) and two mutants with affinities

similar to natural consensus designs. Notably, EA35 N4P/K7P/A8Q/Q49K/N58Y also has strong stability (68 ºC) and acceptable yield (7.0 ± 0.5 mg/L).

### *2.4.5 Cellular Binding Specificity, Protein Oligomerization and Solubility*

To address the possibility of nonspecific binding, A431 (~3 million EGFR per cell) and MCF-7 (< 5,000 EGFR per cell) cell lines were incubated with EA68 and six representative charge-reduced clones at 500 nM and analyzed using flow cytometry (Figure 2-10). For all clones tested, median fluorescence of MCF-7 samples were not statistically different from those containing 0 nM ligand ($p = 0.15$). Additionally, at 500 nM ligand, fluorescent signal was 2 - 3 orders of magnitude higher for A431 cells than MCF-7 in all cases.

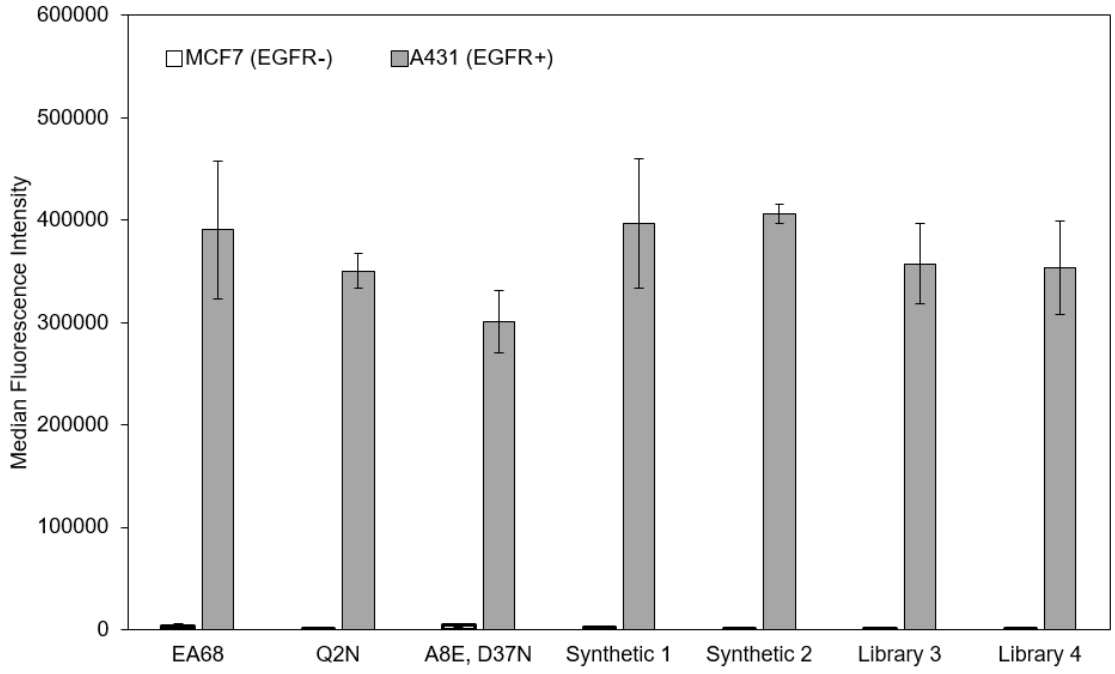



**Figure 2-10. *Cellular EGFR binding specificity*. MCF7 (EGFR-) and A431 (EGFR+) cell lines were incubated on ice with 500 nM affibody ligand, labeled with fluorescein-conjugated anti-$His_6$ antibody, and median fluorescence determined by flow cytometry (n = 3). Bars and error bars represent means and standard error.**

EA68 and seven EA35 affibodies were analyzed with size exclusion chromatography to assess their degree of oligomerization (Figure 2-11). Aprotinin (6.5 kDa) and cytochrome c (12.3 kDa), used as standards, eluted in 15.9 and 13.9 mL retention volumes respectively. EA68 eluted closest to aprotinin at 16.2 mL with all EA35 clones eluting between 17.0 and 20.1 mL of PBS with an average ± standard deviation of 18.0 ± 1.1 mL. Modest peak broadening was also observed. For all proteins tested, no aggregation was visible up to and including concentrations of 1 mg/mL in PBS.

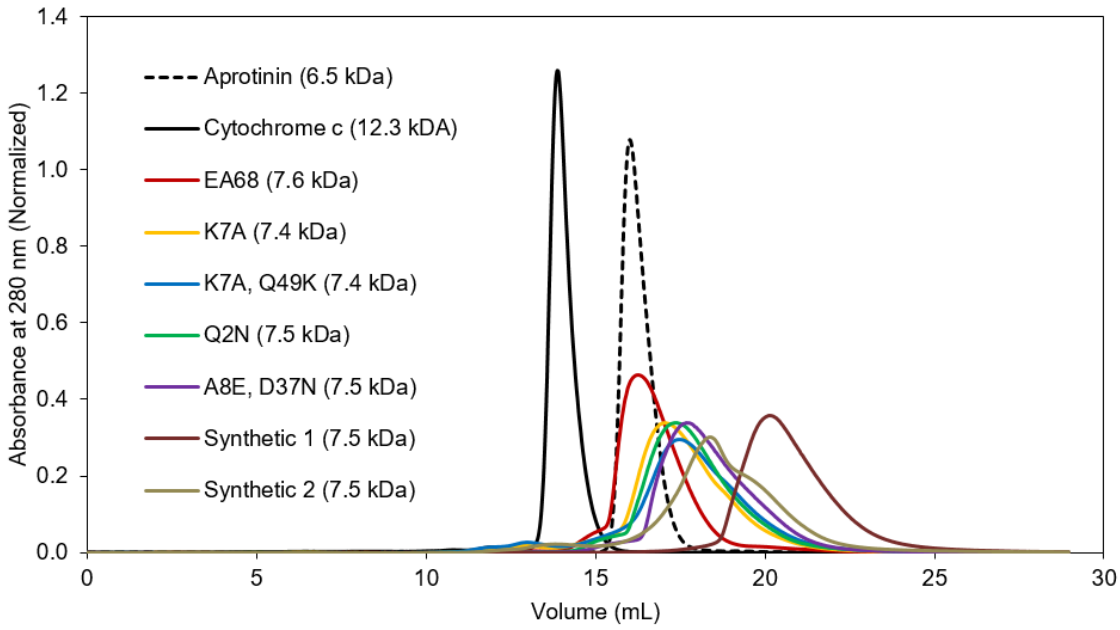



**Figure 2-11. *Size exclusion chromatography of mutant subset.* Lyophilized proteins were resuspended in PBS to 1 mg/mL of which 0.2 mg was passed through a Superdex 75 30/100 size exclusion column at 0.5 mL/min and elution monitored by absorbance at 280 nm. Aprotinin and cytochrome c were used as molecular weight standards.**

## 2.5. Discussion

Consensus design was successful in achieving comparable affinity ($K_d$ = 11 ± 4 nM for EA35 vs. 5 ± 2 nM for EA68) and recombinant yield (4.0 ± 0.8 mg/L vs. 3.5 ± 0.3 mg/L) when neutralizing six charged residues from the parental affibody but exhibited

significantly reduced thermal stability ($T_m$ = 62 ºC vs. 71 ºC) (Figure 2-3). The results reveal that it is valuable to explore various mutants near the consensus design as four of the other ten homolog-guided mutants had higher stabilities, three had nominally higher affinities (albeit with differences of modest statistical power), and eight of ten mutants had higher yields. In particular, EA35 A8E/D37N provides 10 ºC higher thermal stability, 1.6 ± 0.9-fold stronger affinity, and 3.2 ± 0.9-fold increased recombinant yield relative to the consensus mutant EA35. In fact, EA35 A8E/D37N is comparable to the original EA68 despite the neutralization of six charged residues. Thus, while consensus design is a validated approach, it does not always yield the optimal solution. Amino acid frequency in natural homologs can be useful for identifying tolerated amino acids, but their exact choice should not be dictated by the precise natural frequency as evidenced by the lack of correlation between performance (affinity, stability, or yield) and natural frequency for the frequently observed mutations (Figure 2-12).

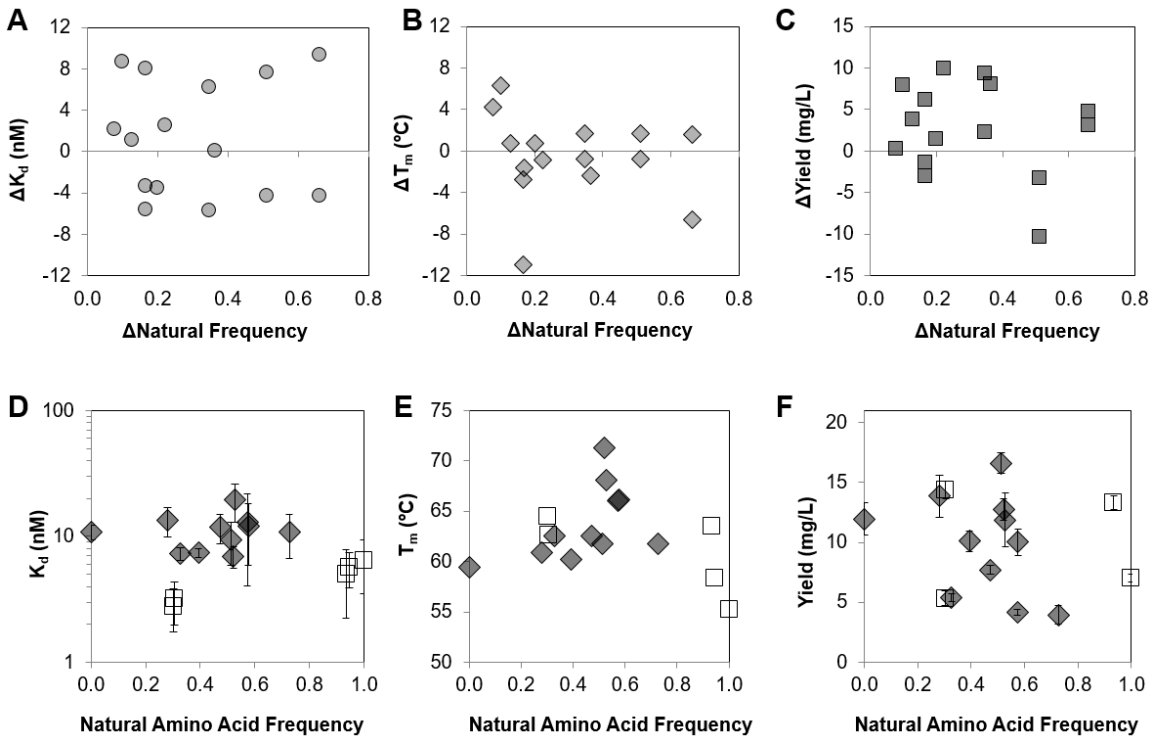



**Figure 2-12.** ***Relationships between performance (affinity, stability, and yield) and amino acid frequency in natural homologs.*** **(A-C) For sets of mutants differing at a single site, the affinity (A), thermal stability (B), and recombinant yield (C) are plotted against the change in natural frequency of the amino acids at the site of interest. (D-F) For EA35 and its equally charged mutants, the affinity (D), thermal stability (E), and recombinant yield (F) are plotted against the normalized sum of natural frequencies at the six sites mutated relative to EA68. $p > 0.05$ for all plotted relationships.**

The analysis of mutants based on homolog frequencies can be broadened using combinatorial library selections provided an efficient assay exists. For example, thermal stability can be enriched by heating yeast-displayed proteins.[177] While cell death occurs in response to the thermal stress, structural integrity is sufficiently maintained to allow for precise discrimination via fluorescence-activated cell sorting; plasmids can then be recovered and amplified by PCR. In addition, affinity selections are straightforward. Libraries were synthesized that include $7 \times 10^5$ mutants using degenerate codons including

amino acids frequent in homologs as well as rare or unrepresented residues. Combinatorial library sorting enabled identification of a mutant with higher affinity than the top 11 consensus mutants (9.2 ± 5.7-fold stronger than EA35) and thermal stability superior to all but one variant (7 ºC higher than EA35). Yet the other three selected mutants were comparably or less effective than the consensus mutants. The ability of the thermal and binding selection pressure to enrich the highly functional mutant from a collection of poorly functional mutants (Figure 2-7A) highlights the power of the approach. Yet, the inability of the selection to deplete some of the modestly effective clones relative to the highly effective clones highlights the limitations of the technique. The selection of comparable affinities to the consensus approach is expected given the selection design: affibodies were incubated with 20 nM EGFR ectodomain to focus selections on stability with retention of good affinity.

Despite the lack of correlation of performance and amino acid frequency in the consensus family of mutants (Figure 2-12), in the combinatorial library selections, highly enriched mutations were only observed for naturally frequent mutations (Figure 2-7B) thereby indicating no dominant false negatives from consensus design. There were some modest false negatives such as E8Q (16% in natural homologs to 43% in selected mutants), K58Y (0% naturally to 27% in selected mutants), and eight other mutations absent naturally but present at least 10% in selected mutants. As in the direct consensus design of the top 11 clones (Figure 2-3), there are significant false positives: K4N (93% naturally to 24% in selected mutants) and K58N (86% naturally to 5% in selected mutants).

Beyond identification of a mutant with improved affinity, the combinatorial library selection – and subsequent sequence analysis of functional mutants – provides an

additional amino acid distribution to guide consensus design. This 'synthetic consensus design' identified a mutant – from only four tested – with 6 ºC higher thermal stability, 6.4 ± 4.3-fold stronger affinity, and 1.8 ± 0.5-fold higher recombinant yield than EA35. In fact, this mutant had higher yield, 3.2 ± 1.9-fold ($p = 0.05$) higher affinity, and comparable stability to parental EA68 despite the neutralization of six charged residues. A second mutant also had high affinity (1.8 ± 0.4 nM), and the other two synthetic mutants had comparable affinities to the natural consensus series resulting in a set of mutants with significantly improved ($p = 0.05$) affinities relative to natural consensus design (Figure 2-9). It is noteworthy, that the top clones from the combinatorial library selection and the synthetic consensus design had two and three mutations (D53Y/K58Y or K4P/K7P/K58Y), respectively, that are rare in natural homologs (0 - 2%). Thus, broadened searches of sequence space were needed to identify such mutants.

This EGFR-binding affibody proved to be amenable to all three modes of engineering charge. The optimal mode for engineering other molecules could be dependent on the molecule of interest as well as the desired properties of interest, but the results herein provide valuable guidance. The synthetic consensus approach proved to be the most effective (highest affinities and comparable stabilities to natural consensus, which were superior to combinatorial library) albeit with modest additional effort relative to natural consensus design. Nevertheless, it is a straightforward approach. Moreover, the synthetic consensus approach is highly valuable in instances in which natural homolog data is not abundant, such as synthetic proteins, or highly conserved at the sites of interest.[168] In cases where natural homolog data is abundant and a rapid, efficient engineering approach is especially valued, natural consensus design can be effective for engineering charge

neutralization. Based on the results with EA68, however, it would be recommended to test several mutants of high natural frequency. This is evidenced by the improvement of several consensus mutants relative to EA35, in particular EA35 A8E/D37N.

In all, several mutants were engineered with a reduced number of charged residues relative to the parental EA68 while maintaining stability, recombinant yield, and affinity (which in some cases was actually improved). Receptor specificity was retained, demonstrated by binding comparisons between low (MCF-7) and high (A431) EGFR-expressing cell lines for a representative subset of clones. All mutants tested exhibited preferential binding to A431 cells two to three orders of magnitudes higher than MCF-7 as represented by median fluorescence. Size exclusion chromatography indicated that clones were monomeric at experimentally relevant concentrations. Retention volumes were higher than predicted by molecular weight alone and peaks were broadened somewhat, but these results may be due to a generally 'stickiness' between affibody and column resin. Taken together such results validate these molecules as useful variants to examine the impact of charge on physiological distribution.

On a different note, it is worth highlighting that the initial experiment in this study examined the ability to merge an evolved framework and an evolved paratope in the affibody domain. A subset of the optimized framework mutations – originally identified in the context of a HER2 binding affibody[144] – were efficiently integrated with a paratope evolved at 13 sites to bind EGFR without detriment to affinity or stability. This, along with a similar result observed with a HER3-binding paratope[178], suggests a modular capability of these particular framework mutations, although paratope- and target-dependent impacts could be expected in some cases.

## 2.6. Conclusions

We were able to apply a selection of the improved affibody framework mutations to an evolved EGFR-binding affibody domain and further mutate the molecule for charged-to-neutral mutations of six residues. The molecule was relatively robust to charge removal yet showed differential retention of stability, affinity, and yield dependent upon the site and residue of mutation. A consensus design approach of 11 charge-reduced clones yielded decreased affinity (median 2.0-fold $K_d$ increase) and stability (median $\Delta T_m$ = 9 °C) with yields exceeding that of parental affibody (median 2.8-fold higher). A broader search of sequence space via a homolog-biased combinatorial library yielded comparable affinities (12 ± 6 nM vs. 11 ± 1 nM), lower stabilities (57 ± 4 ºC vs. 64 ± 1 ºC), and recombinant yields (6.7 ± 2.0 mg/L vs. 9.8 ± 1.2 mg/L). Consensus design from the sequences evolved in the synthetic combinatorial library yielded similar results to the natural consensus design with modestly improved affinity (to 1.7 ± 0.5 nM). In all, the three methods of charge reduction produced a range of phenotypes including several promising lead molecules with strong affinity, stability, and recombinant yield. Moreover, the comparison of charge modulation techniques provides valuable insight for future charge engineering.

## 2.7. Acknowledgements

Experimental assistance by Nicole Olson is appreciated. This work was supported by the University of Minnesota and the National Institutes of Health (R21 EB019518).

## 2.8. Supplemental Figures

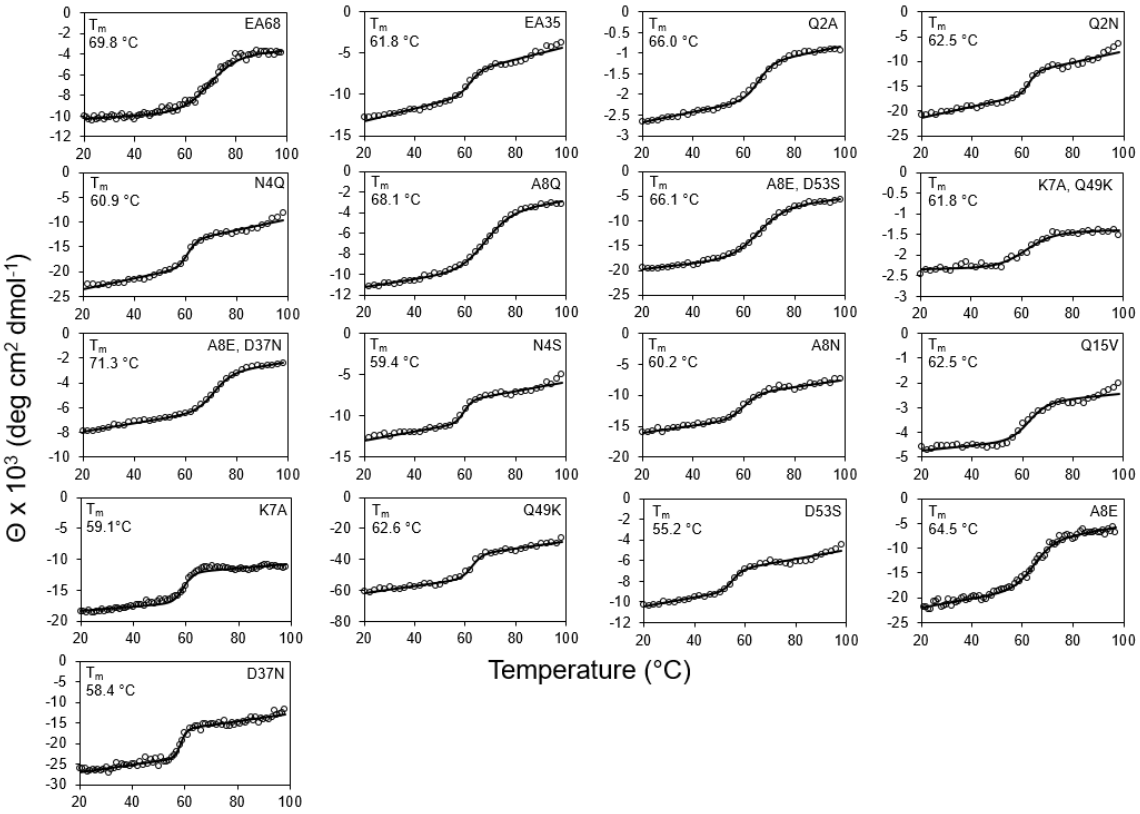



**Figure 2-13.** ***Thermal stability by circular dichroism spectroscopy.*** **Purified affibody domains were analyzed by circular dichroism spectroscopy at 220 nm for temperatures from 20 ºC to 98 ºC. Representative data from one or two analyses are presented.**

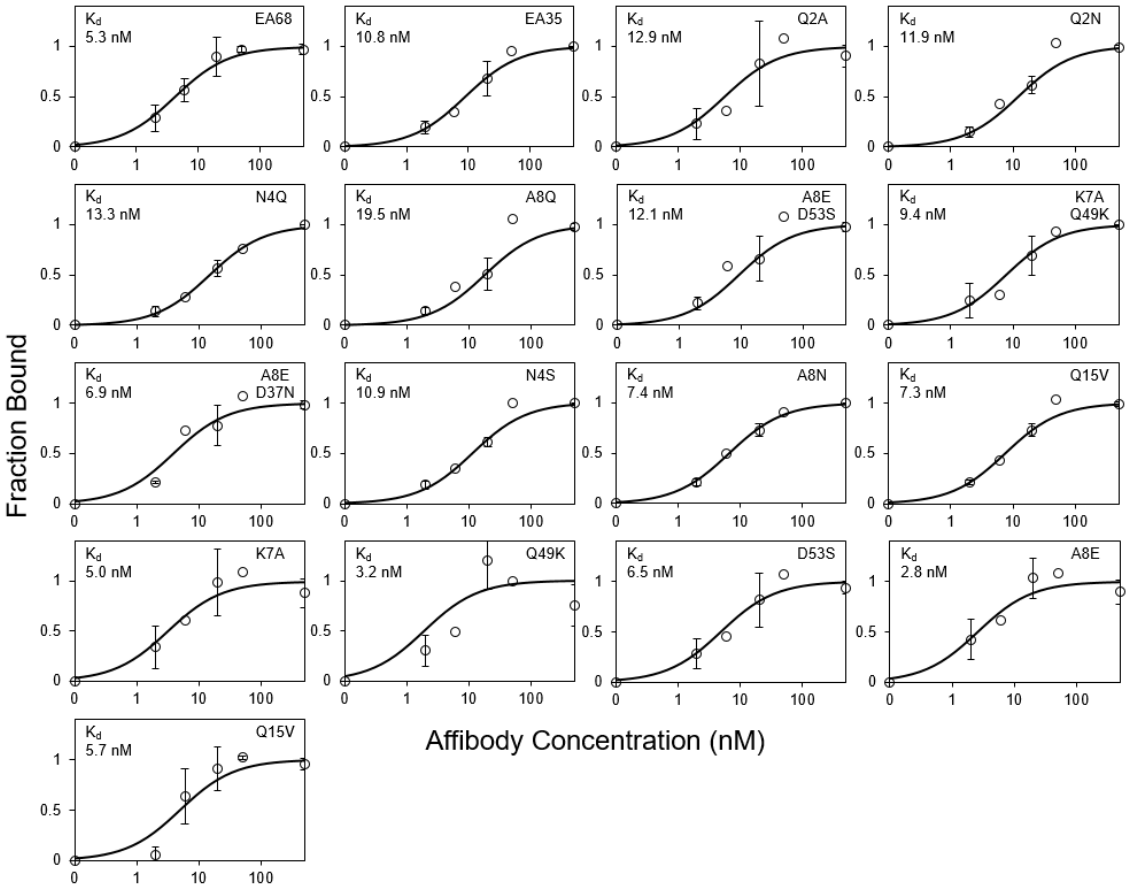



**Figure 2-14.** ***Affinity titration of mutants.*** **Purified affibody domains were added to EGFR-overexpressing A431 cells at the indicated concentration on ice. Binding was detected with fluorescein-conjugated anti-His6 antibody and flow cytometry.**

# Chapter 3 - Evaluation of affibody charge modification identified by synthetic consensus design in molecular PET imaging of epidermal growth factor receptor

## 3.1. Outline

Tumor overexpression of epidermal growth factor receptor (EGFR) correlates to therapeutic response in select patient populations. Thus, molecular positron emission tomography (PET) imaging of EGFR could stratify responders versus nonresponders. We previously demonstrated effectiveness of a "synthetic consensus" design principle to identify six neutralizing mutations within a 58-amino acid EGFR-targeted affibody domain. Herein, we extend the approach to identify additional neutralized variants that vary net charge from -2 to either -4 or +4 while retaining high affinity (1.6 ± 1.2 nM and 2.5 ± 0.7 nM), specific binding to EGFR, secondary structure, and stability ($T_m$ = 68 ºC and 59 ºC). We radiolabeled the resultant collection of five charge variants with $^{64}$Cu and evaluated PET imaging performance in murine models with subcutaneously xenografted $EGFR^{high}$ and $EGFR^{low}$ tumors. All variants exhibited good $EGFR^{high}$ tumor imaging as early as 1 h, with EA35S (+3/-5) achieving 7.7 ± 1.4 %ID/g tumor at 4 h with 1.5 ± 0.3%ID/g $EGFR^{low}$ tumor, 34 ± 5 tumor:muscle and 12 ± 3 tumor:blood ratios. The positively charged EA62S mutant (+6/-2) exhibited 2.2-3.3-fold higher liver signal than the other variants ($p < 0.01$). The EA68 variant with higher charge density was more stable to human and mouse serum than neutralized variants. In a comparison of radiometal chelators, 1,4,7-triazacyclononane,1-glutaric acid-4,7-acetic acid (NODAGA) exhibited superior physiological specificity to 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA). In total, these studies comparatively evaluated a set of EGFR-targeted

affibodies varying in net charge and charge density, which revealed functional variations that are useful in engineering an ideal probe for translational studies.

### 3.2. Introduction

Small protein scaffolds are evolvable towards high target affinity and molecular specificity to create valuable imaging[131,179], therapeutic[180,181], and diagnostic agents[182–184]. Numerous scaffold characteristics influence ligand efficacy including scaffold structure[185], hydrophilicity[38,186], stability [187,188], and ionic charge[38,80,189].

Ligand charge modulation has yielded significant impacts on physiological distribution and retention. Cationic charge has increased renal retention in several cases: cystine knots rich in arginine versus serine[79], affibodies with more lysines in the chelator[80], and cationic synthetic polypeptides versus amphoteric and anionic variants[154]. Other experimental results implicate anionic charge in renal retention: cystine knots rich in glutamic acid versus serine[79], bombesin peptide analogs with increased glutamic acids[82], and affibodies with increased glutamic acids in the chelator[189]. Thus, charge manipulation impacts renal retention with both cationic and anionic residues implicated. Yet charge removal is not systematically sufficient to reduce renal retention. Intermediate charge removal was optimal for some fibronectin domain variants[38]. Comparison of cationic and anionic Fab fragments revealed that increased positive charge corresponded to nearly a 2-fold increase in lung localization whereas elevated negative charge resulted in an almost 3-fold higher stomach signal[84]. Cationic charge sped plasma clearance kinetics[71] for albumin[75], IgG[158,190], Fab[190], and dextran[191]. Cationic charge improved tumor vascular permeability – a key limitation for solid tumor targeting[192] – for albumin[75], IgG[75] and

nanoparticles[78]. Moreover, tumor interstitium is negatively charged[76] and there is some evidence of selective retention of positively charged entities, including liposomes[77], though this could hinder molecular selectivity[73]. In short, ligand charge can be impactful, though the relationships between charge density and net ligand charge with off-target uptake and physiological performance are not well understood, making *a priori* predictions difficult and necessitating several mutants for evaluation. One interpretation of the data suggests a reduction in charge density may aid performance. Yet, it is important to modulate charge without diminishing desirable ligand characteristics including high production yield and soluble recovery, strong affinity, and thermal stability. However, these charge modifications can be particularly difficult in the context of a small ligand[38,193] where such mutations would encompass a greater portion of the protein's structure[98].

A ligand that may benefit from charge modification is the affibody scaffold[194]. Affibodies are 58-residue, three-helix bundles based on the B domain of *staphylococcal* protein A[133] with 13 sites on helices 1 and 2 commonly diversified for binding. They have been used extensively to molecularly target clinically relevant biomarkers such as transferrin[195], TNF-α[196], HER2[174], and gp120[197] with pM to μM affinities. Affibodies are also broadly applicable having been used for molecular imaging[198,199], therapy[200,201], and biotechnology[202,203].

We previously hybridized a known epidermal growth factor receptor (EGFR) targeted paratope region ($Z_{EGFR:1907}$)[160] and an optimized backbone framework containing seven mutations which were incorporated to improve stability and reduce immunoglobulin binding[144]. The resultant affibody ligand, EA68, containing 6 positive and 8 negative residues, was engineered for charge reduction[204]. Consensus design from natural

homologs[162,163,205] and mutation tolerance data generated from high throughput screening (*i.e.* 'synthetic consensus' design) were used to neutralize six charged residues on EA68 creating clones $EA35_{A8E,D37N}$ and Synthetic $2_{N4P,K7P,A8Q,Q49K,N58Y}$ herein known as 'EA35C' and 'EA35S'. Both charge-reduced clones maintained single-digit nanomolar affinity (6.9 ± 1.4 nM, 1.7 ± 0.5 nM), high thermal stability (71 °C, 68 °C), and sufficient recombinant yield (12.7 ± 0.9 mg/L, 7.0 ± 0.5 mg/L).

Epidermal growth factor receptor (EGFR), also known as HER1 or ErbB1, is a glycoprotein in the ErbB family[206] that has been shown to play a key role in tumor survival, growth, and metastasis[107]. EGFR is overexpressed in several tumor types including head and neck, breast, renal, non-small cell lung, colorectal, ovarian, pancreatic, and bladder cancers[207] making it an important prognostic or predictive biomarker and therapeutic target. Surface overexpression and gene copy number are predictors of disease response to receptor (cetuximab) and tyrosine kinase (gefitinib) inhibitor therapies in both primary and metastatic colorectal cancers[110,111,208] as well as non-small-cell lung cancers[112,113,209]. EGFR overexpression alone was associated with poor prognosis for patients with HER2-positive primary breast cancer and predictive of response to trastuzumab therapy[115]. Notably, these results were independent of EGFR copy number and inapplicable to metastatic occurrences.

Thus, effective molecular imaging of EGFR would provide an unmet clinical need and allow for appropriate patient stratification to differentiate responders from nonresponders. Multiple molecular probes have been preclinically developed for positron emission tomography (PET) of EGFR including small engineered proteins – affibodies[117,210–213], fibronectin domains[38,94], and Gp2[214] (and nanobodies for single-photon

emission computed tomography[116,215]) – antibodies[216–220], small molecule inhibitors[221–224], and the native ligand[225,226]. Further development of such probes, to advance sensitivity and selectivity, has potential benefit for EGFR imaging.

Herein, we extend use of mutant tolerance data to explore further charge modulation of EGFR-targeted affibody domains. We evaluate the set of mutants – varied in net charge and charge density – for PET molecular imaging of epidermal growth factor receptor in murine tumor models. Effective imaging is achieved, and functional variations between mutants informs probe development.

## 3.3. Materials and Methods

### *3.3.1 Clone Production and Chelator Conjugation*

Affibody gene construction and ligand production were performed as described previously[204]. EA26S and EA62S genes were created through overlap extension PCR[171], digested using BamHI-HF and NheI-HF restriction enzymes (New England Biolabs, Ipswich, MA), ligated into pET-22b vector containing a C-terminal hexa-histidine (Novagen, EMD Millipore, Billerica, MA) using T4 DNA ligase (New England Biolabs), and transformed into T7 express high efficiency *Escherichia coli* (New England Biolabs) via heat shock. EA68, EA35C, and EA35S plasmids, created previously[204] and stored at -20 °C, were also transformed into T7 express using the heat shock method. Proper transformants were selected on lysogeny broth (LB) based agar plates (15 g/L agar, 10 g/L tryptone, 10.0 g/L sodium chloride, 5.0 g/L yeast extract in $H_2O$) containing kanamycin antibiotic (50 mg/L). After overnight growth at 37 °C, individual colonies were added to 5 mL of liquid LB with kanamycin (50 mg/L) and shaken at 250 rpm, 37 °C until saturated

(8-12 h). Upon saturation, 1 mL of culture was added to 100 mL of sterile LB and shaken at 250 rpm, 37 °C until optical density values at 600 nm reached 0.5-1. Cells were induced with 100 μL of 0.5M isopropyl b-D-1-thiogalactopyranoside (IPTG) for 2 h at 25-28 °C. For instances where larger protein yields were required, production values were scaled to 5 mL saturated culture, added to 1 L LB, and induced with 1 mL of IPTG; all other variables held constant. Production cultures were pelleted, supernatant discarded, and cells resuspended in lysis buffer (50mM sodium phosphate (pH 8.0), 0.5M sodium chloride, 5% glycerol, 5mM 3-[(3-cholamidopropyl)dimethylammonio]-1- propanesulfonate, and 25mM imidazole). After four freeze-thaw cycles, cells were centrifuged at 12,000 g for 12 min, and soluble fraction recovered. Hexa-histidine-tagged affibodies were isolated by immobilized metal affinity chromatography using 0.2 mL HisPur Ni-NTA resin (Thermo Fisher Scientific, Waltham, MA) spin columns (100 mL production culture) or 2 mL resin drip columns (1 L production culture). Protein concentration was determined using absorbance at 280 nm on a Synergy H1 microplate reader (BioTek, Winooski, VT). Large scale productions were further purified by reversed-phase high-pressure liquid chromatography (HPLC) with a C18 column and 15 min gradient of 25% to 80% elution buffer (90% acetonitrile, 9.9% $H_2O$, 0.1% trifluoroacetic acid) with a running buffer of 99.9% $H_2O$, 0.1% trifluoroacetic acid. Clones were snap frozen in liquid nitrogen and lyophilized with a FreeZone Plus$^{2.5}$ freeze-dry system (LabConco, Kansas City, MO). Affibodies used for PET/CT imaging were resuspended in dimethyl sulfoxide, 2-3% trimethylamine v/v, and either 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA)-ester (Macrocyclics Inc., Plano, TX) or 1,4,7-triazacyclononane,1-glutaric acid-4,7-acetic acid (NODAGA)-ester (CheMatech, Dijon, France) at 3:1-10:1 molar ratios of

ester:protein. Upon conjugation, the reaction was quenched with excess 1 M Tris buffer, pH 8.0, isolated via HPLC, freeze-dried, and stored at 4 °C until used experimentally.

### *3.3.2 Mammalian Cell Growth*

A431 epidermoid carcinoma ($2.9 \pm 1.4 \times 10^6$ EGFR/cell[227]), MDA-MB-435 melanoma-type ($1.5 \pm 1.1 \times 10^4$ EGFR/cell[227]), and MCF7 ($1.2 \pm 1.5 \times 10^2$ EGFR/cell[227]) mammary carcinoma cell lines kindly provided by Dr. Daniel Vallera (University of Minnesota), Dr. Tim Starr (University of Minnesota), and Dr. Deepali Sachdev (University of Minnesota), respectively, were cultured in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% fetal bovine serum and 1% penicillin-streptomycin at 37 °C, 5% $CO_2$.

### *3.3.3 Affinity Measurement, Target Specificity, and Circular Dichroism*

At 70-80% confluence, culture media was removed and 0.25% trypsin-EDTA in Hank's Balanced Salt Solution (HBSS) (Corning, Tewksbury, MA) added for 6-10 min to detach cells. Cells were washed once with cold PBSA (phosphate buffered saline: 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, and 1 g/L bovine serum albumin) to prevent EGFR internalization. Approximately 40,000 A431 or MCF7 cells were labeled with each affibody at varying concentrations (0-500 nM for affinity titrations and 500 nM for specificity tests) at 4 °C. Cells were again washed in cold PBSA and incubated with fluorescein-conjugated rabbit anti-His6 antibody (ab1206, Abcam, Cambridge, MA). Median fluorescence of each sample was determined with an Accuri C6 (BD Biosciences, San Jose, CA) and affinity values calculated by least squares regression assuming a 1:1

binding event between ligand and receptor. Equilibrium dissociation constants are presented as the average ± standard deviation of titrations.

Lyophilized affibodies were resuspended to 0.5-1 g/L in 1xPBS. 300 μL of each suspension was added to a quartz cuvette with 1 mm path length and ellipticity measured from 200 to 260 nm at 20 °C on a Jasco J-815 spectrophotometer. Thermal denaturation curves were created by monitoring ellipticity at 220 nm as the sample was heated from 20 to 98 °C at a rate of 2 °C/min. The midpoint of denaturation ($T_m$) was determined using a two-state protein unfolding model. After heating, suspensions were cooled to 20 °C and scanned from 200 to 260 nm to ascertain the degree to which each affibody refolds.

### *3.3.4 Tumor Xenotransplantation*

Six-week-old female ($Foxn1^{nu}/Foxn1^{nu}$) mice (Jackson Labs, Bar Harbor, ME) were anesthetized with 2.0% isoflurane in 0.8-1 mL/min oxygen and subcutaneously injected in their right shoulder with approximately 10 million MDA-MB-435 cells in a 50:50 v/v Matrigel Matrix/DMEM suspension. After 4.5 weeks, this method was repeated with approximately 4 million A431 cells injected subcutaneously into right shoulders. At six weeks after the MDA-MB-435 inoculation, xenografts from both tumor lines had grown to 4-11 mm in diameter. All animal studies were approved by the University of Minnesota Institutional Animal Care and Use Committee (Protocol 1411-32007A).

### *3.3.5 $^{64}$Cu Chelation of DOTA or NODAGA/Purification*

10-12 mCi of $^{64}CuCl_2$ (University of Wisconsin) was suspended in 100 μL of 100 mM NaAc, pH 5.0 for buffering and titrated to pH 6.0 using 0.1 M HCl and 0.1 M NaOH (final volume: 150-225 μL). 1-2 mCi of buffered $^{64}CuCl_2$ was added to 500 μL of DOTA-

affibody or 500 μL of NODAGA-affibody at 70-120 μM in PBS pH 7.4 or 10 mM sodium acetate pH 6.0 respectively and placed at 37 °C for 1 h to elicit complexation. $^{64}$Cu-DOTA-affibody and $^{64}$Cu-NODAGA-affibody were separated from free copper using a PD-10 size exclusion column and PBS or 10 mM sodium acetate elution buffer.

### *3.3.6 RadioTLC/Enzymatic Degradation*

$^{64}$Cu-NODAGA-affibody purity levels and enzymatic degradation were ascertained with an AR-2000 radio-thin layer chromatography scanner (Eckert and Ziegler, Berlin, Germany). For determination of purity, 2 μL of PD-10 elution volume 3.5-4.0 mL was spotted on filter paper and a PBS mobile phase was used for migration. Area under the curve analysis was used to determine the fraction of unconjugated copper found in the solvent front peak compared to the relatively immobile $^{64}$Cu-NODAGA-affibodies. One μL from each $^{64}$Cu-NODAGA-affibody PD-10 elution volume 3.5-4.0 mL was added to 20 μL of normal human serum (Jackson ImmunoResearch), normal mouse serum (Jackson ImmunoResearch), or 0.25% trypsin-EDTA in HBSS and incubated at 37 °C for 4 hours. Two μL samples were taken from each mixture at time points 1 and 4 h, spotted on filter paper (PBS mobile phase), and scanned to determine the relative percentage of affibody metabolites by means of area under the curve integration.

### *3.3.7 PET Imaging/Quantification*

Mice were anesthetized under 2% isoflurane at 1 mL/min oxygen until no longer ambulant. Approximately 2.2 to 3.7 MBq of $^{64}$Cu-NODAGA- or $^{64}$Cu-DOTA-affibody, as determined by an Atomlab 100 dosimeter (Biodex, Shirley, NY) calibrated for $^{64}$Cu

measurements, was injected through lateral tail veins. Five-minute PET scans were performed at 1 and 4 h post-injection on an Inveon microPET/CT (Siemens, Malvern, PA) and reconstructed with 2D ordered-subset expectation maximization (OSEM2D). Four iterations of Fourier rebinning were performed. Photopeak energy cutoffs were set to 350-650 keV with a time window of 3.438 ns. PET images were smoothed with a 1x1x1 voxel Gaussian filter. CT scans were executed using 340 projections of 80kV at 500 μA with 4x4 binning and an effective pixel size of 85.6 μm. Scans ran for approximately 384 s and were reconstructed with the Feldkamp-David-Kress algorithm with slight noise reduction and Shepp-Logan filtering.

PET image quantification was performed with Inveon Research Workplace software (Siemens). CT images were used as guides to determine 3D regional limits to tissues of interest including the liver, kidneys, thigh muscle, A431, and MDA-MB-435 tumors. The maximum average PET signal in a 3x3x3 spherical voxel (9.1 $mm^3$) was determined for each tissue type.

### *3.3.8 Tissue Quantification*

After imaging, mice were euthanized by cervical dislocation while still under isoflurane anesthesia. After verifying death via toe pinch, mice were dissected and blood, heart, lung, liver, pancreas, spleen, stomach, intestine, skin, A431 tumor, MDA-MB-435 tumor, brain, bone, muscle, and kidneys were collected and weighed. Sample radioactivity was measured by either a CRC 25W (Capintec, Florham Park, NJ) gamma counter over all energy ranges or an Automatic Gamma Counter (Hidex, Turku, Finland) between 400 and

600 keV with all reads averaged over 60 s. Counts from both instruments were calibrated using serial dilutions based on the dose reported by the Atom Lab 100 dosimeter.

#### *3.3.9 Statistics*

Comparisons for statistical significance were performed using two-tailed student's t-tests assuming unequal variance. Values are reported as mean ± standard error unless otherwise stated.

### 3.4. Results and Discussion

#### *3.4.1 Further charge modulation of EA35S -- guided by sequence tolerance data from high throughput selections -- is well tolerated with regards to secondary structure, stability, affinity, and recombinant yield.*

In a previous study[204], 3 basic and 3 acidic residues from the EGFR-targeted affibody EA68 were neutralized via mutation yielding 19 mutants of which clones 'EA35S' and 'EA35C' retained the greatest combination of target affinity, thermal stability, and recombinant yield. While these clones enable the evaluation of charge density reduction on physiological distribution, additional variations of net charge were desired. Based on the success of the 'synthetic consensus' strategy for identifying mutants with three basic and five acidic residues (+3/-5)[204] the approach was used to identify alternative mutations. The most frequently observed neutralizing mutations – within the previously enriched affibody mutants – were chosen to mutate EA35S to create clones with +2/-6 (EA26S) and +6/-2 (EA62S) charge (Figure 3-1) with the exception of position 47 in EA62S. Although not tested in the synthetic approach, site 47 was mutated to glutamine as necessary based

on the prevalence of accepted glutamic acid to glutamate replacements at other sites, E14 residing in the paratope, and residue 36 remaining negatively charged (glutamic of aspartic acid) in 98% of natural homologs.

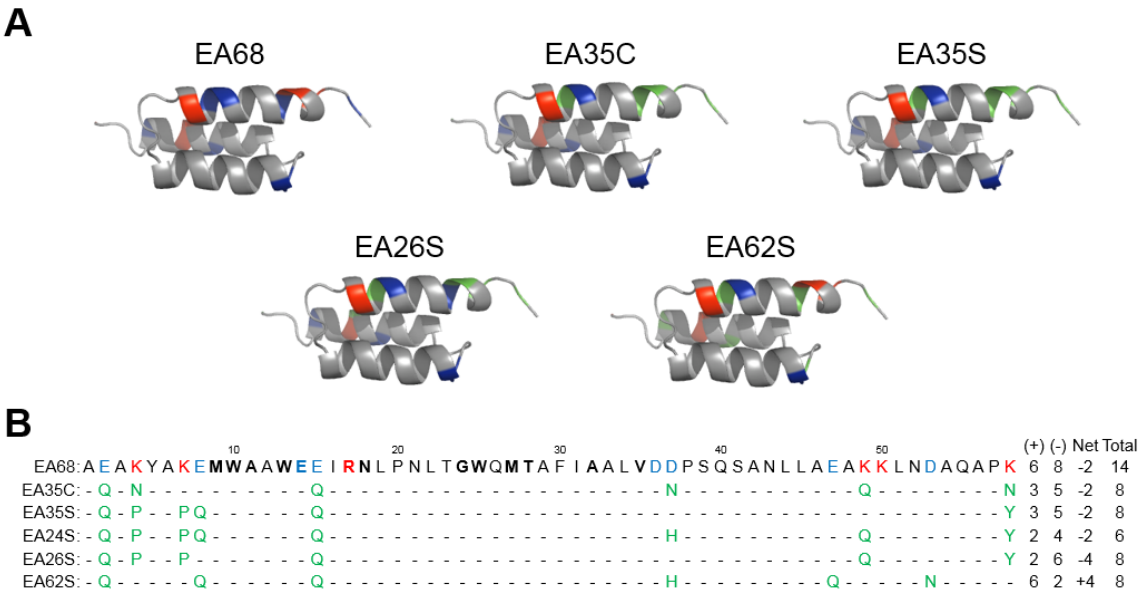



**Figure 3-1. *Affibody mutants.* (A) Structural schematics of affibody clones EA68, EA35C, EA35S, EA26S, and EA62S. Acidic, basic, and charged-to-neutral mutated residues are designated as red, blue, and green, respectively. Structures are based on PDB 1H0T[133]. (B) Sequences of the affibody mutants with the same color scheme as in (A). The traditional paratope regions are shown in bold.**

EA26S and EA62S were readily produced and purified from the soluble fraction of *E. coli* culture (0.62 ± 0.02 mg/L and 1.9 ± 0.3 mg/L, respectively). The purified proteins were titrated for binding to EGFR-overexpressing A431 cells and evaluated by flow cytometry. EA26S and EA62S bind with high affinity (1.6 ± 1.2 nM and 2.5 ± 0.7 nM, respectively; Figure 3-2A) comparable to EA35S (1.7 ± 0.5 nM) and exceeding the EGFR affinities of both EA68 (5.3 ± 1.7 nM) and EA35C (6.9 ± 1.4 nM). Binding is molecularly specific as demonstrated by the lack of binding to $EGFR^{low}$ MCF7 cells (Figure 3-2B).

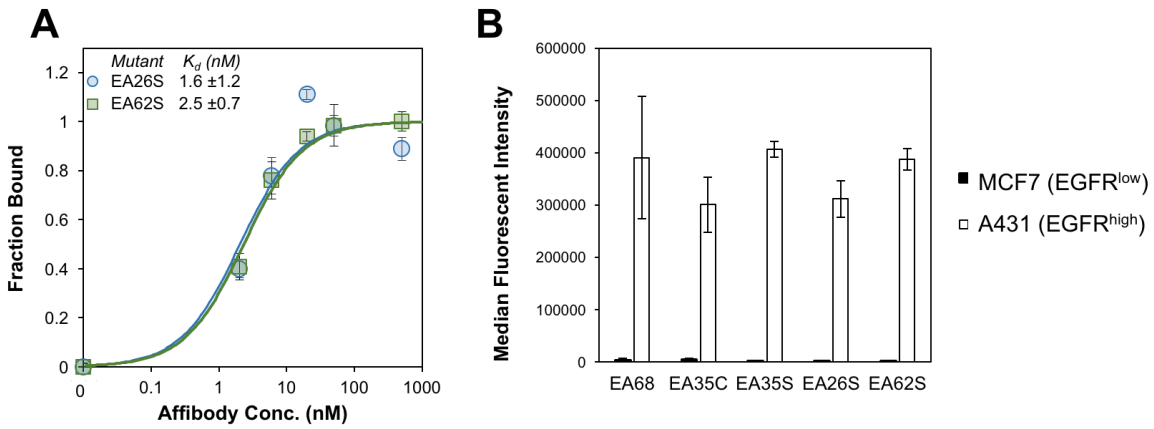



**Figure 3-2. *Binding characterization.* (A) EGFR$^{high}$ A431 cells were mixed with the indicated concentration of affibody (EA26S (blue circles) or EA62S (green squares)). Binding was detected with fluorophore-conjugated anti-His6 antibody via flow cytometry. Equilibrium dissociation constants were calculated assuming a 1:1 binding model. n = 3 measurements per variant. (B) EGFR$^{high}$ A431 cells (white) or EGFR$^{low}$ MCR7 cells (black) were incubated with 500 nM affibody. Binding was detected as in (A). n = 3 measurements per variant.**

Like EA35S and EA35C, the new variants exhibit comparable secondary structure and ability to refold after thermal denaturation to the parental EA68 (Figure 3-3A,B). EA26S exhibits thermal stability ($T_m$ = 68 ºC) comparable to EA68, EA35C, and EA35S (71, 71, and 68 ºC, respectively) whereas EA62S is less stable ($T_m$ = 59 ºC) (Figure 3-3C).

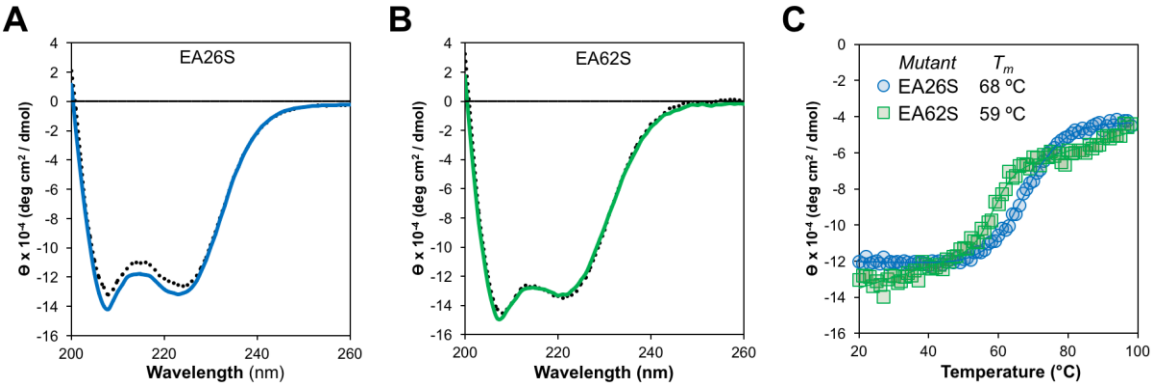



**Figure 3-3. *Structural characterization.* Purified EA26S and EA62S were evaluated by circular dichroism. (A,B) Wavelength scans from 200 nm to 260 nm before (solid line) or after (dotted line) heating to 98 ºC and cooling to 20 ºC. (C) Thermal scan from 20 ºC to 98 ºC at a wavelength of 220 nm.**

Thus, the collection of EGFR-binding affibodies exhibits similar biophysical properties but varied charge distribution (Table 3-1).

**Table 3-1. Biophysical properties of affibody clones used for PET/CT imaging.**

| Variant | Yield (mg/L) | $K_d$ (nM) | $T_m$ (°C) | 2° Structure | Specificity (A431:MCF7 Binding Signals) |
|---|---|---|---|---|---|
| EA68 | 3.5 ± 0.3 | 5.3 ± 1.7 | 71 | α-helical | 109 ± 63 |
| EA26S | 0.62 ± 0.02 | 1.6 ± 1.2 | 68 | α-helical | 214 ± 134 |
| EA62S | 1.9 ± 0.3 | 2.5 ± 0.7 | 59 | α-helical | 342 ± 209 |
| EA35S | 7.0 ± 0.5 | 1.7 ± 0.5 | 68 | α-helical | 529 ± 297 |
| EA35C | 12.7 ± 0.9 | 6.9 ± 1.4 | 71 | α-helical | 74 ± 29 |

The efficacy of consensus design based on synthetic combinatorial library selections is consistent with previous efforts.[168,204] This efficacy may also be supported by the robustness of the previously improved affibody framework.[144] The resultant ability to modulate charge yielded a set of affibody variants that enable evaluation of the physiological effects of these mutations, which include changes in net charge and charge density.

### *3.4.2 Affibody variants are functional EGFR imaging probes*

The parental EA68 and the four neutralized charge variants with a range of net charge were conjugated with 2,2′-(7-(1-carboxy-4-((2,5-dioxopyrrolidin-1-yl)oxy)-4-oxobutyl)-1,4,7-triazonane-1,4-diyl)diacetic acid, the N-hydroxysuccinimidyl ester of NODAGA. Conjugation was verified by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Figure 3-4) and showed an average labeling of 1.7, 1.7, 1.5, 0.9, and 1.4 NODAGA per ligand for EA68, EA35C, EA35S, EA26S, and EA62S, respectively.

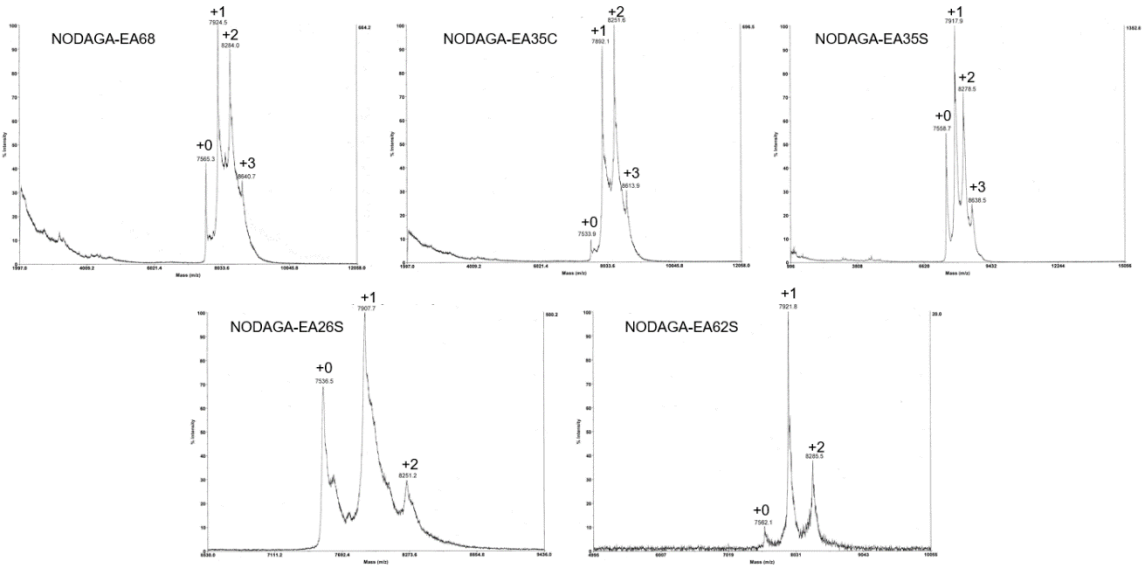



**Figure 3-4.** ***Conjugation of NODAGA to affibody charge variants.*** **Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) of NODAGA-affibody charge variants with 0, 1, 2, or 3 degrees of conjugation**

While more site-selective conjugation could have been achieved with unique cysteine incorporation or alternative approaches, the amine-conjugated affibodies exhibited similar target affinities to their unconjugated forms (Figure 3-5).

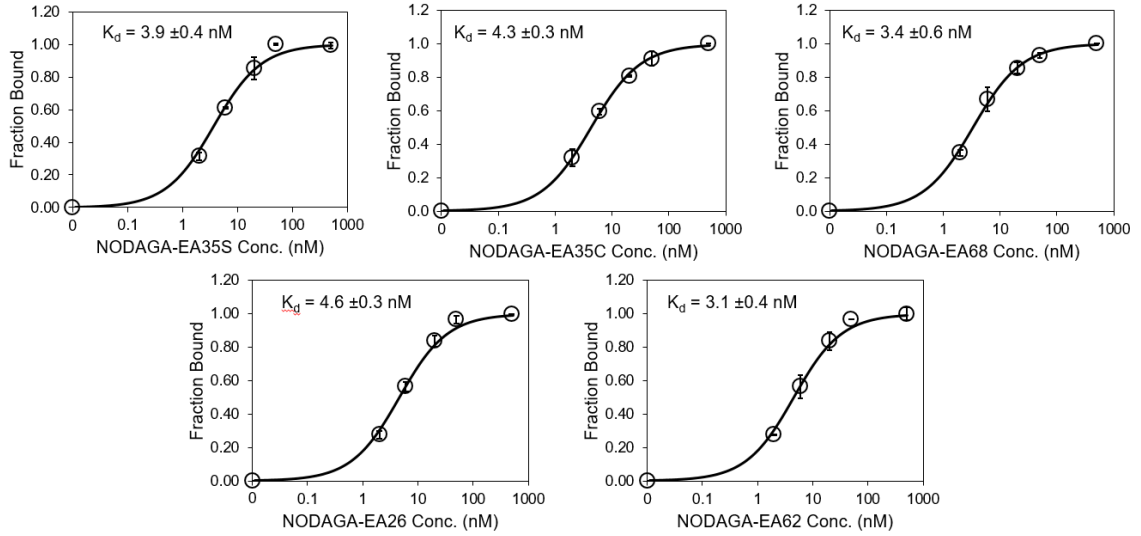



**Figure 3-5.** ***NODAGA-affibody target-affinity characterization.*** **EGFR-overexpressing A431 cells were mixed with the indicated concentration of NODAGA-conjugated**

**affibody. Binding was detected with fluorophore-conjugated anti-His6 antibody via flow cytometry. Equilibrium dissociation constants were calculated assuming a 1:1 binding model. n = 3 measurements per variant. NODAGA-conjugated versions of EA35S and EA26 were nominally weaker binders to EGFR than their unconjugated counterparts ($K_d$ = 3.9 ± 0.4 nM vs. 1.7 ± 0.5 nM and 4.6 ± 0.3 vs. 1.6 ± 1.2 nM respectively). However, these variances are slight and should not meaningfully impact *in vivo* biomarker localization. EA68 and EA62 showed no difference in binding affinity upon NODAGA conjugation and EA35C presented a slight improvement (4.3 ± 0.3 nM vs. 6.9 ± 1.4 nM).**

Each NODAGA-conjugated affibody was mixed with 1 - 1.5 mCi of $^{64}$Cu in 100 mM sodium acetate buffer at pH 6.0 to undergo complexation at 37 °C for 1 h. Radiolabeled proteins were separated from free $^{64}$Cu through size exclusion chromatography, yielding high radiochemical purity (96 - 99%) and labeling efficiency (65 ± 6%) for subsequent murine injections. By applying knowledge of previous purifications of their non-radiolabeled forms, approximate specific activities were calculated between 0.4 - 1.1 MBq/nmol.

Imaging performance and physiological distribution were evaluated in mice harboring dual xenografted tumors: EGFR$^{high}$ A431 and EGFR$^{low}$ MDA-MB-435. $^{64}$Cu-complexed affibodies were injected via tail vein, and PET/computed tomography (CT) images were collected at 1 and 4 h post-injection. CT scans were used as locational guides to determine regions of interest including the liver, kidneys, muscle (anterior/posterior thigh), A431, and MDA-MB-435 tumors. All affibody variants exhibited strong EGFR imaging performance at both 1 h and 4 h (Figures 3-6 and 3-7).

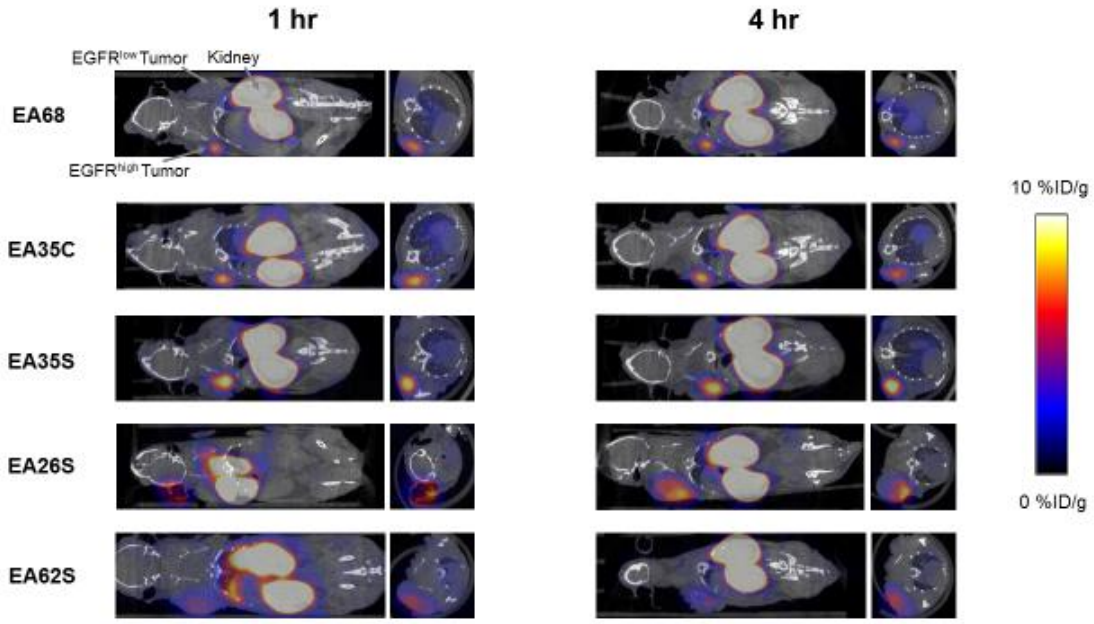



**Figure 3-6. *PET/CT imaging.* Mice with xenografted A431 (EGFR$^{high}$) and MDA-MB-435 (EGFR$^{low}$) tumors in either shoulder were injected intravenously with $^{64}$Cu-NODAGA-affibody (five different variants) and imaged via PET/CT at 1 h and 4 h post-injection. Coronal and axial slices through the tumor are shown for representative mice (n = 3 - 7).**

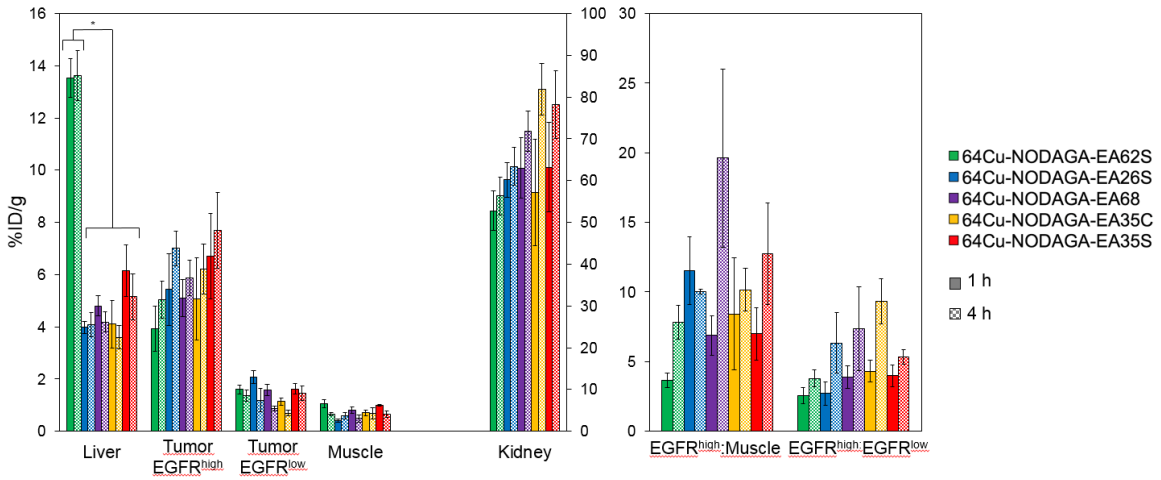



**Figure 3-7. *PET imaging quantification.* Mice (n = 3 - 7) with xenografted A431 (EGFR$^{high}$) and MDA-MB-435 (EGFR$^{low}$) tumors in either shoulder were injected intravenously with $^{64}$Cu-NODAGA-affibody (five different variants) and imaged via PET/CT at 1 h (solid) and 4 h (checkered) post-injection. Signal in the liver, tumors,**

**muscle, and kidney were quantified using Siemens Inveon Research Workplace software. * designates $p < 0.01$.**

At 1 h, the five affibodies had localized to EGFR-overexpressing A431 tumors with molecular specificity (5.2 ± 0.5 %ID/g) and cleared from $EGFR^{low}$ MDA-MB-435 tumors (1.6 ± 0.1 %ID/g). By 4 h, A431 signal had increased further (6.7 ± 0.5 %ID/g) while MDA-MB-435 remained low and decreased slightly (1.2 ± 0.1 %ID/g). Among clones, EA35S displayed the nominally highest EGFR-binding at both 1 h (6.7 ± 1.6 %ID/g) and 4 h (7.7 ± 1.4 %ID/g). As is common with renal clearance of small proteins, high levels of kidney signal were prevalent, and average concentration exceeded 50 %ID/g for all clones. Note that due to the large number of photons emanating from the kidneys, multiple coincident events may be detected within the allotted time window forcing some events to be discarded to avoid locational ambiguity. This can result in kidney activity measurements being artificially lower than their true values and cross-comparison among clones may not be valid[228]. Background clearance was rapid with an average muscle signal of 0.8 ± 0.1 %ID/g at 1 h which decreased to 0.6 ± 0.1 %ID/g by 4 h. Excluding EA62S, clones exhibited moderate levels of liver signal due to hepatic processing at both 1 h (4.8 ± 0.3 %ID/g) and 4 h (4.6 ± 0.3 %ID/g). Conversely, EA62S had significantly greater hepatic retention than all other clones at both 1 h (13.5 ± 0.7 %ID/g, $p < 0.01$) and 4 h (13.6 ± 0.9 %ID/g, $p < 0.01$).

Imaging results were corroborated by analysis of excised tissues at 4 h post-injection (Figure 3-8). The parental protein, EA68, exhibited strong EGFR-specific tumor uptake (7.2 ± 1.5 %ID/g in $EGFR^{high}$ versus 0.9 ± 0.2 %ID/g in $EGFR^{low}$; $p = 0.002$; Figure 3-8A). Specificity was further supported by 11 ± 4 tumor:blood and 126 ± 56 tumor:muscle

ratios (Figure 3-8B). Retention is substantial in kidneys (169 ± 12 %ID/g) and modest in liver (4.5 ± 0.6 %ID/g). Similarly, EA35C displayed elevated $EGFR^{high}$ (7.0 ± 0.9 %ID/g) and low $EGFR^{low}$ tumor (1.5 ± 0.4 %ID/g) signal with substantial tumor:blood (11 ± 4 %ID/g) and tumor:muscle (151 ± 70 %ID/g) ratios. EA35S showed the highest level of A431 retention (12 ± 2 %ID/g) but a nominally elevated hepatic signal (7 ± 1 %ID/g) versus clones EA35C, EA26S, and EA68 which averaged 4.4 ± 0.3 %ID/g in the liver. Compared to all other variants, which are negatively charged, positively charged EA62S has significantly higher liver and spleen retentions of 14 ± 1 %ID/g ($p \leq 0.01$) and 3.2 ± 0.3 %ID/g ($p \leq 0.01$) respectively.

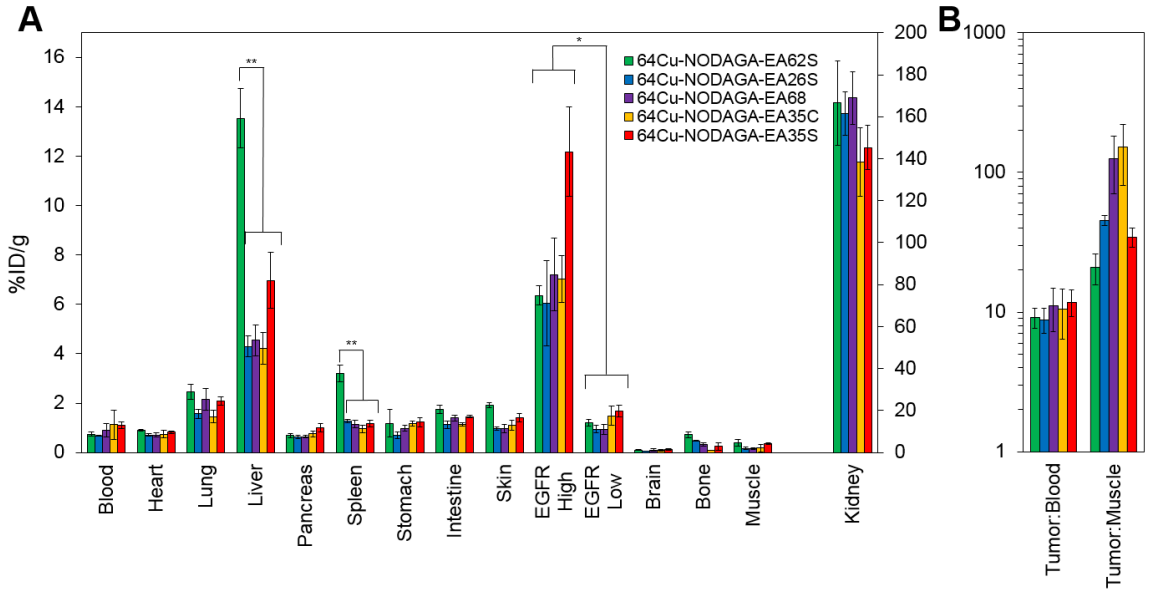



**Figure 3-8.** ***Excised tissue analysis.*** **Mice with xenografted A431 ($EGFR^{high}$) and MDA-MB-435 ($EGFR^{low}$) tumors in either shoulder were injected intravenously with $^{64}$Cu-NODAGA-affibody (five different variants) and euthanized at 4 h post-injection. Tissues or fluid was excised and weighed; radioactivity was measured (n = 3-7). **: $p < 0.01$, *: $p < 0.05$.**

The rapid distribution of the affibody probes, as well as clearance of unbound probe, enables effective imaging at early time points. While not requisite for diagnostic imaging, additional studies could be used to evaluate probe localization at later times.

### *3.4.3 Chelator Comparison*

To determine the effect that chelator may have on biodistribution, EA68 and EA35S were conjugated with the DOTA by reaction of N-hydroxysuccinimidyl-DOTA ester with primary amines on the N-terminus or lysine residues of each protein. MALDI-MS was used to verify an average labeling of 1.1 DOTA per EA68 and 2.0 DOTA per EA35S affibody (Figure 3-9).

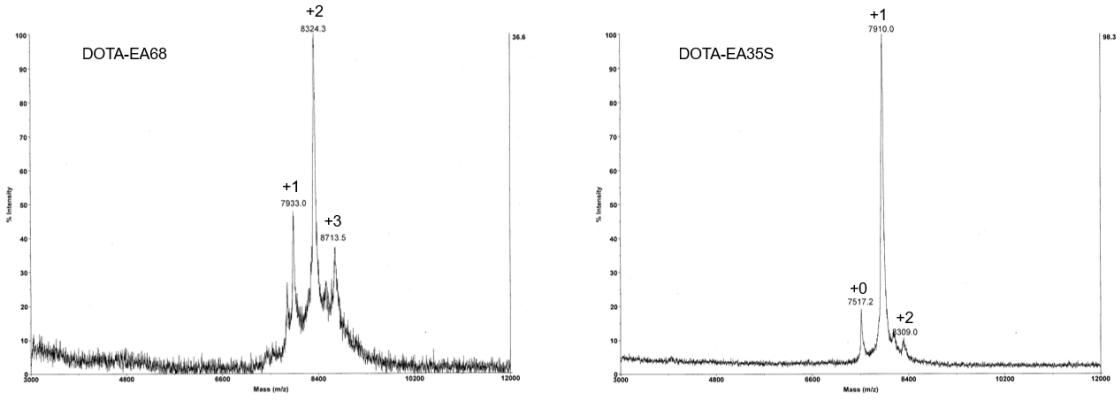



**Figure 3-9. *Conjugation of DOTA to EA68 and EA35S.* Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) of DOTA-EA68 and DOTA-EA35S with 0, 1, 2, or 3 degrees of conjugation**

DOTA labeled EA68 and EA35S were mixed with 1.9 - 2.0 mCi of $^{64}$Cu in PBS at pH 7.4 to undergo complexation at 37 °C for 1 h. Size exclusion chromatography was used to separate complexed affibodies from free $^{64}$Cu providing a labeling efficiency of 43% and 60% for $^{64}$Cu-NODAGA-EA35S and $^{64}$Cu-NODAGA-EA68. The approximate specific activities of EA35S and EA68 were 0.25 and 0.33 MBq/nmol.

Chelator performance was evaluated in six nude mice with human EGFR$^{high}$ A431 and EGFR$^{low}$ MDA-MB-435 tumors xenografts. A431 xenotransplantation and subsequent growth was successful in all murine models. However, four of six mice showed no measurable MDA-MB-435 tumor progression. The remaining two mice displaying both cancer lines were split to provide one negative control tumor for each DOTA-conjugated affibody. Mice were injected intravenously through lateral tail veins with 2.3 - 3.7 MBq of $^{64}$Cu-NODAGA-EA35S and $^{64}$Cu-NODAGA-EA68. At 4 h post-injection mice were euthanized, tissues excised, and blood collected.

The use of a NODAGA chelator yields lower signal in all background (i.e. non-A431 tumor) tissues in comparison to DOTA (Figure 3-10). $^{64}$Cu-NODAGA-EA35S maintained significantly lower blood, heart, liver, and stomach uptake than $^{64}$Cu-DOTA-EA35S. Similarly, $^{64}$Cu-NODAGA-EA68 had substantially lower signal in the lungs and intestine than $^{64}$Cu-DOTA-EA68. EGFR$^{high}$ A431 tumor localization was comparable (and nominally greater) for both NODAGA-labeled EA35S and EA68 than their DOTA-conjugated counterparts. These results are consistent with previous data indicating imperfect stability of the $^{64}$Cu-DOTA complex *in vivo*.[229]

As numerous EGFR PET probes are in preclinical development, performance comparison is desirable. Strict quantitative comparisons have moderate utility because of differences in tumor models and experimental procedures but can provide guidance when such differences are limited and acknowledged. The current $^{64}$Cu-NODAGA-EA35S exhibits comparable tumor uptake and tumor:muscle specificity but superior tumor:blood (12 ± 3 vs. 1.0 ± 0.1 without blocking) and liver specificity (5.2 ± 0.8 %ID/g vs. 19 ± 3 %ID/g) relative to a Miao et al. study[230] with $^{64}$Cu-DOTA-affibody$_{1907}$. The chelator

comparison experiments (Figure 3-10) reveal that this benefit derives predominantly from the use of NODAGA as the $^{64}$Cu-DOTA-affibody results are comparable between the current study and Miao et al. and similarly matched by an additional study with a Hep3B xenograft.[231] Other alternative radiochemistries, with $^{89}$Zr-DFO[Ref. 212] or $^{18}$F-CBT[Ref. 213], also reduce hepatic signal albeit without benefit to tumor:blood specificity. Strong tumor:blood specificity was achieved with the $^{64}$Cu-DOTA-affibody via a small dose of unlabeled affibody.[230] The previous $^{64}$Cu-DOTA-affibody$_{1907}$ and the current $^{64}$Cu-DOTA-EA68 only differ in the current use of the engineered protein framework, optimized by Feldwisch et al.[144] for enhanced stability and reduced immunoglobulin binding. Thus, by historical comparison, these mutations enable these benefits without substantially altering physiological delivery; yet direct comparison in the same experimental system would be needed to discern potential differences more precisely. The current $^{64}$Cu-NODAGA-affibodies yield comparable PET imaging and physiological delivery performance to a $^{64}$Cu-DOTA-Gp2 probe[214] and higher tumor sensitivity and tumor:muscle specificity relative to a $^{64}$Cu-DOTA-fibronectin[94].

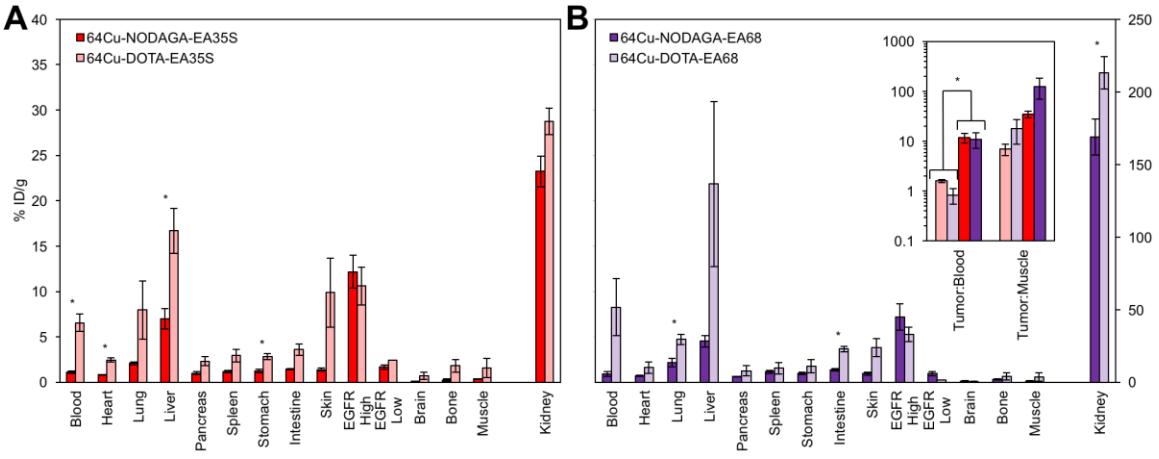


**Figure 3-10.** ***Chelator comparison.*** **Mice with xenografted A431 ($EGFR^{high}$) and MDA-MB-435 ($EGFR^{low}$) tumors in either shoulder were injected intravenously with $^{64}$Cu-NODAGA-affibody or $^{64}$Cu-DOTA-affibody and euthanized at 4 h post-injection. Tissues or fluid was excised and weighed; radioactivity was measured (n = 3 - 7). * designates $p < 0.05$.**

### *3.4.4 Labeling and Serum Stability*

*In vitro* enzymatic stability was assessed through radio-thin-layer chromatography. Radiolabeled charge variants were added to murine serum, human serum, or trypsin-EDTA (0.25%) in HBSS at a 1:20 volumetric ratio and incubated for 4 hours at 37 °C. Area under the curve analysis shows that parental EA68 was most resistant to degradation in both human and murine serum to similar degrees. Approximately 2% and 4% of sample signal contained $^{64}$Cu-labeled metabolites at both 1 and 4 h in human and murine serum, respectively (Figure 3-11). Similarly, EA35C and EA26S were stable in human serum with 6% and 4% of total signal area due to proteolysis after 4 hours. EA35S and EA62S degraded more rapidly, with 10% and 11% of cleaved $^{64}$Cu-peptide in human serum at 4 h. Excluding EA68, clones were more susceptible to enzymatic cleavage in murine serum than human serum. This difference was most pronounced in clones EA26S and EA62S, where 16% and 44% of radioactivity was attributed to metabolic product in mouse serum versus 4% and 11% for human serum. In all cases, trypsin digestion resulted in hydrolysis products displaying the furthest positional shift indicating nearly complete proteolysis.

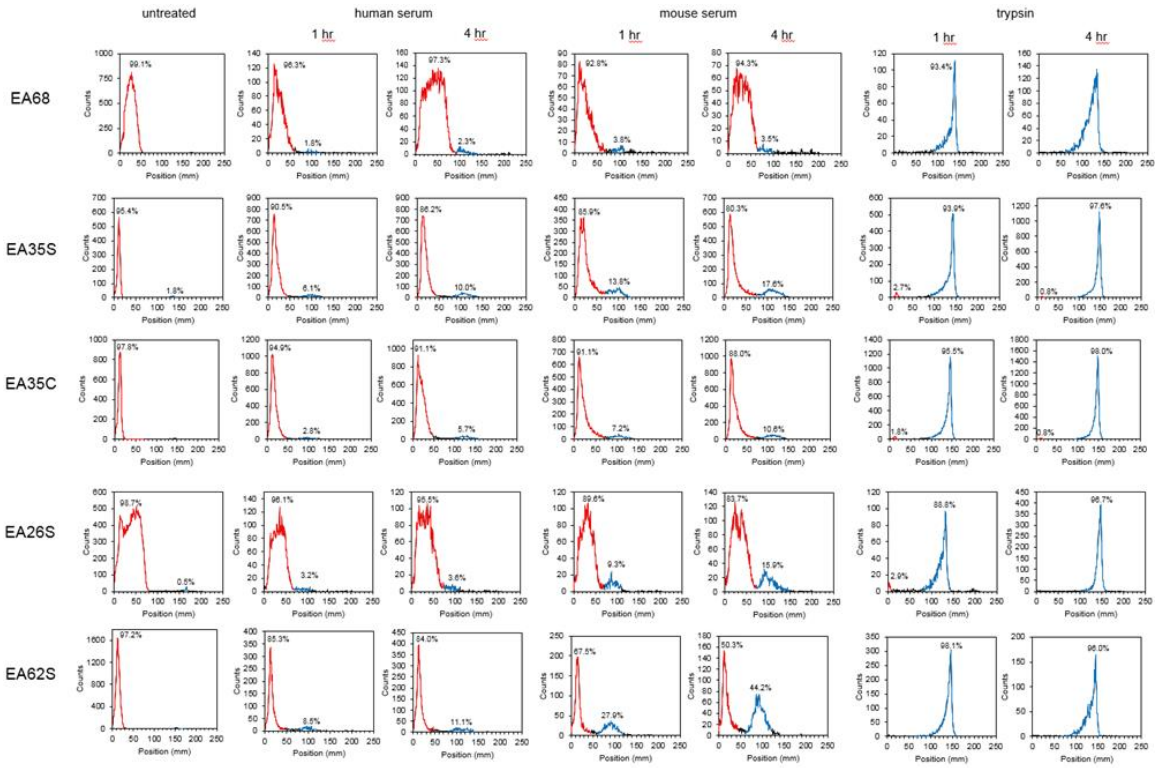



**Figure 3-11. *Proteolytic stability.* Radio chromatography of affibody clones left untreated, in serum, or in trypsin.**

### 3.5. Conclusions

Molecular imaging of EGFR is poised to address an important clinical need: to stratify a subset of cancer patients as responders vs. nonresponders to several molecularly targeted therapies to enable effective treatment planning. While numerous probes have undergone preclinical development, further molecular advances are needed as no clinical imaging probe is yet available for EGFR. Charge has been shown to play an important role in physiological distribution of proteins, but system complexity precludes *a priori* design and necessitates experimental evaluation. We have further demonstrated the efficacy of 'synthetic consensus design' in identifying functional neutralizing mutations to yield a set of small affibody scaffolds varied in net charge and charge density. Evaluation of this set

of probes, radiolabeled with $^{64}$Cu, for PET imaging of EGFR in mouse tumor models reveals sensitive, specific molecular imaging for multiple variants with notable differences in liver retention, specificity, and probe stability for different charge and chelator designs. Elucidation of these differential efficacies aids engineering of an ideal probe for translational studies.

### 3.6. Acknowledgements

This work was supported by the National Institutes of Health (R21 EB021511 to B.J.H.) and the University of Minnesota's Interdisciplinary Dissertation Fellowship (to M.A.K.).

# Chapter 4 - Engineered charge redistribution of Gp2 proteins through guided diversity for improved PET imaging of epidermal growth factor receptor

## 4.1. Outline

The Gp2 domain is a protein scaffold for synthetic ligand engineering. However, the native protein function results in a heterogeneous distribution of charge on the conserved surface, which may hinder further development and utility. We aim to modulate charge, without diminishing function, which is challenging in small proteins where each mutation is a significant fraction of protein structure. We constructed rationally guided combinatorial libraries with charge-neutralizing or charge-flipping mutations and sorted them, via yeast display and flow cytometry, for stability and target binding. Deep sequencing of functional variants revealed effective mutations both in clone-dependent contexts and broadly across binders to epidermal growth factor receptor (EGFR), insulin receptor, and immunoglobulin G. Functional mutants averaged 4.3 charge neutralizing mutations per domain while maintaining net negative charge. We evolved an EGFR-targeted Gp2 mutant that reduced charge density by 33%, maintained net charge, and improved charge distribution homogeneity while elevating thermal stability ($T_m$ = 87 ± 1 $^oC$), improving binding specificity, and maintaining affinity ($K_d$ = 8.8 ± 0.6 nM). This molecule was conjugated with 1,4,7-triazacyclononane,1-glutaric acid-4,7-acetic acid for $^{64}Cu$ chelation and evaluated for physiological distribution in mice with xenografted A431 ($EGFR^{high}$) and MDA-MB-435 ($EGFR^{low}$) tumors. Excised tissue gamma counting and positron emission tomography / computed tomography imaging revealed good $EGFR^{high}$ tumor signal (4.7 ± 0.5 %ID/g) at 2 h post-injection and molecular specificity evidenced

by low uptake in $EGFR^{low}$ tumors (0.6 ± 0.1 %ID/g, significantly lower than for non-charge-modified Gp2, $p = 0.01$). These results provide charge mutations for an improved Gp2 framework, validate an effective approach to charge engineering, and advance performance of physiological EGFR targeting for molecular imaging.

## 4.2. Introduction

Small protein ligands that bind targets with high affinity and specificity are valuable resources for imaging[131,198], therapeutic[232–234], and diagnostic purposes[182,183]. Small scaffolds benefit from increased tumor extravasation[36,235,236], tissue penetration[192,237] and plasma clearance[36,238] resulting in improved target-to-background contrast at early time points[128–130]. Protein ligands must be engineered for not only binding function but also physicochemical robustness and physiological performance. To optimize performance, several influencing factors have been studied including scaffold structure[185], hydrophilicity[38,186], stability[187,239], and charge[38,80,189]. Protein charge is of considerable interest as clusters of positively charged residues can reduce solubility[67] and drive nonspecific binding to native negatively charged cell membranes[68–70]. An increase in the fraction of charged residues on nonparatope regions can correspond to decreased target affinity[74]. Ligand charge can also affect transvascular flux[75,78], tumor uptake[77], plasma clearance[71,158], and renal retention[38,79]. It is therefore important to be able to modulate charge without diminishing desirable ligand characteristics including high production yield and soluble recovery, strong affinity, and stability. However, these charge modifications can be particularly difficult in the context of a small ligand[38,193] where such mutations would encompass a greater portion of the protein's structure[98].

A scaffold that may benefit from redistributed and reduced charge is the Gp2 domain[149]. Gp2 is a small, stable 45-residue scaffold with α-helical and β-strand structure whose solvent-exposed loop regions have been diversified to discover and evolve binders to numerous targets.[149] Engineered Gp2 domains have been validated for molecular positron emission tomography (PET) imaging of epidermal growth factor receptor (EGFR)[214] and antagonism of insulin receptor (InsR) in tamoxifen-resistant, triple-negative breast cancer cell culture[240]. The native Gp2 framework contains three positive and six negative residues, with each charge set regionally clustered. It is hypothesized that this charge distribution evolved in wild-type Gp2 because of its native function. In a Gp2-RNA polymerase model structure, the string of acidic residues projects into the DNA-binding channel thereby aiding inhibitory function.[134] Also, the basic residues R42 and R44 were experimentally shown as necessary for Gp2 function as they aid binding to RNA polymerase.[134] Yet maintenance of these charges is not necessary – and we hypothesize not optimal – for its newfound role as a scaffold for synthetic ligand engineering. These highly charged regions may drive nonspecific binding[73], off-target retention[38,79,80,82], and decrease protein yield during ligand production due to reduced solubility[151–153]. Therefore, the ability to mutate such clusters without reducing Gp2 affinity or stability has considerable value.

Numerous strategies exist for charge engineering including computational design[151,241,242], homology-guided consensus design[162], and combinatorial library selection[204]. We merged versions of each of these techniques to identify charge-modifying mutations that are well tolerated in multiple Gp2 domains and applied effective mutations

to an EGFR-binding Gp2 domain for PET imaging. EGFR overexpression is present in numerous cancers[161], is correlated with differentiation, and is an independent prognostic indicator of poor disease-free and overall survival in colorectal cancer patients[243,244]. Molecular imaging of EGFR could empower patient stratification of responders versus nonresponders to molecularly targeted therapy.[110,115]

Combinatorial libraries of Gp2 mutants targeting EGFR, InsR, and rabbit immunoglobulin G (rIgG), guided by computational stability[245] as well as natural[163,164] and chemical homologs[246], were sorted via yeast display and flow cytometry for mutants that retained thermal stability and target affinity. Functional mutants elucidated preferred net and total charge states, determined favored amino acids at charged positions, and identified several promising charge-modified anti-EGFR Gp2 ligands. A charge-reduced clone, GαE35, exhibited reduced nonspecific binding and highly functional performance in PET/CT imaging of $EGFR^+$ tumors in murine models. The results advance Gp2 scaffold engineering, as well as charge engineering more broadly, and improve preclinical performance of an EGFR PET agent.

## 4.3. Materials and Methods

### *4.3.1 FoldX and Homolog Consensus*

FoldX[245] was used to predict destabilizing effects of residue mutations at charged positions by estimating the change in free energy of folding for these mutations on the Gp2 scaffold. Calculations were performed at each position using 40 randomized paratopes, or until < 0.1 kcal $mol^{-1}$ absolute or < 1% relative deviation in mean was achieved with each additional variant for four solved crystal structures of T7 phage Gp2 (2LMC[247], 2WNM[134],

4LLG[248], 4LK0[248]) and averaged to determine the final $\Delta\Delta G_{folding,mutation}$ of each canonical amino acid at all positions.

We searched for homologs of the wild-type Gp2 protein via NCBI BLAST[249] (in 2016) resulting in 32 sequences, which were aligned in ClustalW2[250]. Sitewise amino acid frequencies were calculated.

### *4.3.2 Library Construction*

Degenerate oligonucleotides containing the mixed bases necessary for desired residue mutations at charged positions and encoding for $G\alpha E_{2.2.3}$, $G\alpha I_{2.2.5}$, or $G\alpha R_{3.2.3}$ paratope regions were purchased (IDT DNA, Coralville, IA) and assembled via PCR overlap extension. Each of the three libraries was transformed into the yeast surface display strain EBY100[167] using homologous recombination with a linearized pCT vector as described[172]. Transformation efficiency was determined by serial dilution on SD-CAA/agar plates (0.07 M sodium citrate (pH 5.3), 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 20 g/L glucose, and 15-18 g/L bacto agar in water). Sanger and Illumina sequencing of each library was performed to verify proper transformation.

### *4.3.3 Binder Selection and Illumina Sequencing*

Yeast were grown in SD-CAA (0.07 M sodium citrate (pH 5.3), 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, and 20 g/L glucose in water) and induced in SG-CAA (0.1M sodium phosphate (pH 6.0), 6.7 g/L yeast nitrogen base, 5 g/L casamino acids, 19 g/L galactose, and 1 g/L glucose in water) at 30 ºC, 250 rpm as described.[172] Populations with sample sizes between 10x and 20x of theoretical diversity were taken from naïve libraries and sorted using flow cytometry. Yeast were washed in phosphate-buffered saline

with 1 g/L bovine serum albumin (PBSA) heated to 50-80 °C for 30 min, cooled on ice to 4 °C, incubated with 2 – 20 nM recombinant, biotinylated target protein and anti-c-Myc antibody (Cat# 626802, BioLegend, San Diego, CA) to isolate full-length Gp2 mutants and account for differences in surface expression, labeled with streptavidin-Alexa Fluor 647 (Life Technologies, Waltham, MA) and goat anti-mouse-FITC antibody (Cat# F0257, Sigma–Aldrich, St. Louis, MO), then sorted.

Because yeast cell viability diminishes significantly above 42 °C[251], the recovered populations in heat-treated sorts always exceeded $10^5$ yeast cells. This number of cells provided sufficient plasmid reclamation for the resulting downstream plasmid recovery via Zymoprep protocol (Zymo Research, Irvine, CA), PCR amplification of Gp2 genes, and retransformation through electroporation into yeast to rebuild display libraries for continued sorting. After their final recombinant target sorts, Gp2 libraries were exposed to avidin-coated magnetic beads (Invitrogen, Waltham, MA), followed by beads with immobilized nontarget mouse IgG; nonbinding yeast were collected and sequenced.

The EGFR-binding population underwent two additional rounds of sorting without heating against 0.2 and 2 nM solubilized EGFR from A431 cell lysate as described previously.[252] Briefly, after reaching 70-80% confluency, A431 cells were detached, washed in ice-cold PBS, incubated with NHS-$PEG_4$-Biotin, lysed and centrifuged, and the resulting supernatant mixed with Gp2 expressing yeast recovered from bead depletion. Binding was detected via flow cytometry as for recombinant target selections.

Illumina MiSeq paired-end sequencing was used to obtain $4.7 \times 10^5$, $1.6 \times 10^4$, $5.9 \times 10^4$, and $8.9 \times 10^5$, $4.9 \times 10^4$, $5.9 \times 10^4$ reads from the naïve and sorted binding populations of EGFR, InsR, and rIgG Gp2 libraries. Forward and reverse reads were aligned using

PANDAseq[253] with a 90% quality cutoff resulting in 2704, 2840, 595, and 3202, 797, 171 unique sequences from naïve and enriched populations in EGFR, InsR, and rIgG libraries. Sequences were sorted, aligned, and sitewise frequencies calculated. Clones appearing more than once had frequencies dampened by $\sqrt{n}$ where n represents the number of clonal occurrences found in the Illumina sequences.

### *4.3.4 Mammalian Cell Growth*

MDA-MB-435 melanocyte ($(1.5 \pm 1.1) \times 10^4$ EGFR per cell[227]) and A431 epidermoid carcinoma ($(2.9 \pm 1.4) \times 10^6$ EGFR per cell[227]) cell lines were kindly provided by Dr. Tim Starr (University of Minnesota) and Dr. Daniel Vallera (University of Minnesota). Both cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) containing 1% penicillin streptomycin and 10% fetal bovine serum at 37 °C in humidified air with 5% $CO_2$. When needed experimentally, cultures were incubated with 1-2 mL of 0.25% Trypsin-EDTA for 6-10 min to detach cells.

### *4.3.5 GαE Clonal Production and Characterization*

Gp2-encoding regions in DNA recovered from the final lysate-extracted EGFR flow cytometry sort were amplified by polymerase chain reaction, digested with NheI-HF and BamHI-HF restriction enzymes (New England Biolabs, Ipswich, MA), and ligated into a pET-22b vector containing a C-terminal hexa-histidine (Novagen, EMD Millipore, Billerica, MA) with T4 DNA ligase (New England Biolabs). Plasmids were transformed via heat-shock into T7 Express Competent *E. coli* (New England Biolabs) and proper transformants selected on lysogeny broth (LB) plates containing 50 mg/L kanamycin. Surviving colonies were added to 4 mL liquid LB/kanamycin (50 mg/L) and incubated in

shake flasks at 37 °C, 250 rpm for 10-16 h. Saturated cultures were added to 100 mL of LB, incubated 37 °C, 250 rpm until reaching an optical density between 0.5 and 1.0 upon which 0.1 mL 0.5 mM isopropyl β-D-1-thiogalactopyranoside was added to induce protein expression. Induced cultures were incubated 4-24 h at 30 °C, 250 rpm, cells were pelleted and resuspended in lysis buffer (50mM sodium phosphate (pH 8.0), 0.5M sodium chloride, 5% glycerol, 5mM 3-[(3-cholamidopropyl)dimethylammonio]-1- propanesulfonate, and 25mM imidazole), frozen and thawed 4 times, centrifuged for 10 min at 10 °C, and the resulting cell lysate run through 0.2 mL Cobalt HisPur resin volume spin columns (Thermo Fisher Scientific, Waltham, MA) to recover Gp2 proteins. When greater levels of Gp2 production were required, as in the case of murine PET/CT imaging studies, saturated cultures were added to 1 L LB and clones were recovered using 2 mL Cobalt HisPur resin volume gravity columns. Recovered yields were assessed using absorbance at 280 nm on a Synergy H1 microplate reader (BioTek, Winooski, VT).

In some cases, clones were additionally purified by reversed-phase high-performance liquid chromatography with a C18 column using a 15 min gradient of 10% to 90% elution buffer (90% acetonitrile, 9.9% water, 0.1% trifluoroacetic acid) and the remaining solution composed of running buffer (99.9% water, 0.1% trifluoroacetic acid). Recovered Gp2 was lyophilized.

Secondary structure, midpoint of thermal denaturation ($T_m$), and protein refolding after heating was determined using circular dichroism performed on a Jasco J-815 spectrophotometer with samples placed in a 1 mm path length quartz cuvette. Purified and lyophilized Gp2 ligands were resuspended in PBS to 0.2-1 mg/mL. Ellipticity was measured between 205 and 260 nm wavelengths at 20 °C before and after heating to 98 °C.

Thermal denaturation was evaluated at 220 nm during sample heating at 2 °C/min from 20 °C to 98 °C.

At 70-80% confluence, A431 were detached as described previously, pelleted, and resuspended in 4 °C PBSA. Each Gp2 ligand was added at varying concentrations (0, 2, 20, 50, and 500 nM) to 50,000 resuspended A431 cells and incubated at 4 °C to allow for EGFR binding. Cells were again pelleted and washed with cold PBSA, then incubated with fluorescein-conjugated anti-His6 antibody (Abcam, Cambridge, MA). Cells were washed a final time and analyzed for fluorescence using flow cytometry on an Accuri C6 (BD Biosciences, Franklin Lakes, NJ). Using median fluorescence at each ligand concentration, the equilibrium dissociation constant, $K_d$, was determined by minimizing the sum squared errors assuming a 1:1 binding interaction. Affinity titrations were performed thrice. To elucidate EGFR specificity, an analogous procedure was performed with A431 and MDA-MB-435 cell lines at 500 nM target (n = 3).

Gp2 ligands used for PET/CT imaging were resuspended in dimethyl sulfoxide, 2-3% trimethylamine v/v, and 1,4,7-triazacyclononane,1- glutaric acid-4,7-acetic acid (NODAGA)-ester (CheMatech, Dijon, France) at 5:1-10:1 molar ratios of ester:protein. Upon conjugation as verified by matrix-assisted laser desorption ionization mass spectrometry, the reaction was quenched with excess 1 M Tris buffer, pH 8.0, isolated via HPLC, freeze-dried, and stored at 4 °C until use.

### *4.3.6 Tumor Xenotransplantation*

Six-week-old female ($Foxn1^{nu}/Foxn1^{nu}$) mice (Charles River Laboratories, Wilmington, MA) were anesthetized via inhalation of 2% isoflurane in 1-1.2 mL $min^{-1}$ $O_2$.

Ten million MDA-MB-435 cells, suspended in a 50:50 v/v DMEM/Matrigel Matrix solution, were subcutaneously injected in their right shoulder. After 4.5-5 weeks, this process was repeated with 7 million A431 cells subcutaneously injected into the opposing shoulder. Xenograft tumors were grown to 5-12 mm in diameter (approximately 6 weeks after MDA-MB-435 inoculation). In preblocking experiments, this procedure was repeated with 7 million A431 cells subcutaneously injected into both shoulders followed by 1.5 weeks of tumor growth until tumor diameters reached 7-10 mm.

### *4.3.7 Chelation and Purification*

5-12 mCi $^{64}CuCl_2$ (University of Wisconsin-Madison) was diluted in 100 µL 100 mM sodium acetate pH 6.0 to provide buffering, then titrated with 1 M sodium hydroxide and 1 M hydrochloric acid to a final pH of 6.0 as confirmed by narrow band pH paper (final volume: 200-250 µL). 1-2 mCi buffered $^{64}CuCl_2$ was added to 50-120 µM NODAGA-Gp2 in 10 mM sodium acetate pH 6.0 and incubated at 42 °C for 1 h to enact complexation. $^{64}$Cu-NODAGA-Gp2 ligands were separated from free $^{64}$Cu through size exclusion chromatography with PD-10 columns pre-equilibrated with 10 mM sodium acetate pH 6.0.

### *4.3.8 RadioTLC/Enzymatic Degradation*

$^{64}$Cu-NODAGA-Gp2 purity and enzymatic stability were determined with an AR-2000 radio-thin-layer chromatography scanner (Eckert and Ziegler, Berlin, Germany) as described previously[254]. Briefly, 1-2 µL PD-10 elution volume 3.5-4.0 mL was spotted on chromatography paper with PBS used as the mobile phase to separate any uncomplexed copper from $^{64}$Cu-NODAGA-Gp2. Through integration analysis of the resulting position

vs. scintillation count data, relative fractions of free $^{64}$Cu and $^{64}$Cu-NODAGA-Gp2 were determined.

To ascertain enzymatic stability, 1 μL from $^{64}$Cu-NODAGA-Gp2 PD-10 elution volume 3.5 – 4.0 mL was added to 20 μL normal human serum (Jackson ImmunoResearch), normal mouse serum (Jackson ImmunoResearch), or 0.25% trypsin-EDTA in Hank's balanced salt solution and incubated at 37 °C for 4 h. Two μL samples were taken at 1 and 4 h, spotted on chromatography paper with a PBS mobile phase, and analyzed for metabolite migration. Integration analysis was again used to determine the relative percentage of Gp2 metabolites.

### *4.3.9 Tissue Quantification of $^{64}$Cu-NODAGA-GαE22, -GαE35, and -GαE57 in Mice Bearing A431 and MDA-MB-435 Tumor Xenografts*

Mice were anesthetized with 2% isoflurane at 1 mL min$^{-1}$ oxygen until stationary. When no longer ambulatory, 1.0-1.9 MBq $^{64}$Cu-NODAGA-GαE22, -GαE35, or -GαE57 was injected via lateral tail vein. At 2 h p.i., mice were again anesthetized then euthanized by cervical dislocation. After confirming expiration by toe pinch, mice were dissected and blood, heart, lung, liver, pancreas, spleen, stomach, intestine, skin, brain, bone, muscle, kidney, A431, and MDA-MB-435 were extracted and weighed. Resected samples were measured for radioactivity by a CRC-25W (Capintec, Florham Park, NJ) gamma counter across all energy ranges and averaged over 1-2 min. Incident counts were calibrated using serial dilutions based on the dose reported by an Atom Lab dosimeter.

### *4.3.10 PET Imaging/Quantification and Tissue Quantification of $^{64}$Cu-NODAGA-GαE35 in Mice Bearing Dual A431 Tumor Xenografts*

Mice were anesthetized as described previously. Upon loss of motor function, mice bearing dual A431 tumors were injected via lateral tail vein with 0 or 0.7 – 0.8 mg NODAGA-GαE35 followed 5-10 min later by 1.0-2.7 MBq $^{64}$Cu-NODAGA-GαE35. Injected radioactivity was quantified by an Atomlab 100 dosimeter (Biodex, Shirley, NY) calibrated for $^{64}$Cu measurements. Static PET scans were performed at 1 and 2 h post-injection on an Inveon microPET/CT (Siemens, Malvern, PA) and images reconstructed with the 2D ordered-subset expectation maximization (OSEM2D) method using four iterations of Fourier rebinning. Photopeak energy cutoffs were 350-650 keV with a timing window of 3.437 ns. A 1 x 1 x 1 voxel Gaussian filter was used to smooth PET images. Concomitant CT scans were completed using 340 projections of 80 kV at 500 µA with 4 x 4 binning, 200 ms exposure over 384 s total scan time, and an effective pixel size of 85.6 µm. CT scans were reconstructed with the Feldkamp–David–Kress algorithm using slight noise reduction and Shepp-Logan filtering. PET image ROI quantification was performed on Inveon Research Workplace software (Siemens) using CT images as guides to determine 3D margins for tissues of interest including liver, thigh muscle, kidneys, and A431 tumors. The maximum average PET signal in a 3 × 3 × 3 spherical voxel (9.1 $mm^3$) was calculated for each tissue of interest.

Two hours after $^{64}$Cu-NODAGA-GαE35 injection, mice were euthanized, tissues extracted, weighed, and radioactivity measured as described previously.

### *4.3.11 Statistics*

Comparisons for statistical significance between was determined using two-tailed Student's *t*-tests assuming unequal variance between sample data sets. Data are presented as mean ± standard error unless otherwise noted.

## 4.4. Results and Discussion

### *4.4.1 Mutant Library Design and Construction*

To identify functional mutants of the EGFR-binding Gp2 ($G\alpha E_{2.2.3}$, also termed GαE57 since it is a *G*p2 ligand targeting *E*GFR with *five* basic and *seven* acidic residues) with modified charge distributions, we constructed combinatorial libraries for high-throughput functional screening in which the ten charged sites present in the framework were diversified. We also aimed to identify charge mutations that are broadly acceptable within numerous engineered variants of the Gp2 scaffold (Figure 4-1A); thus we created additional libraries for insulin receptor and rabbit IgG binding based on previously established $G\alpha I_{2.2.5}$[150] and $G\alpha R_{3.2.3}$[149] ligands. In all three cases, charged residues in the evolved loops were conserved while ten framework residues (Figure 4-1B) were mutated.

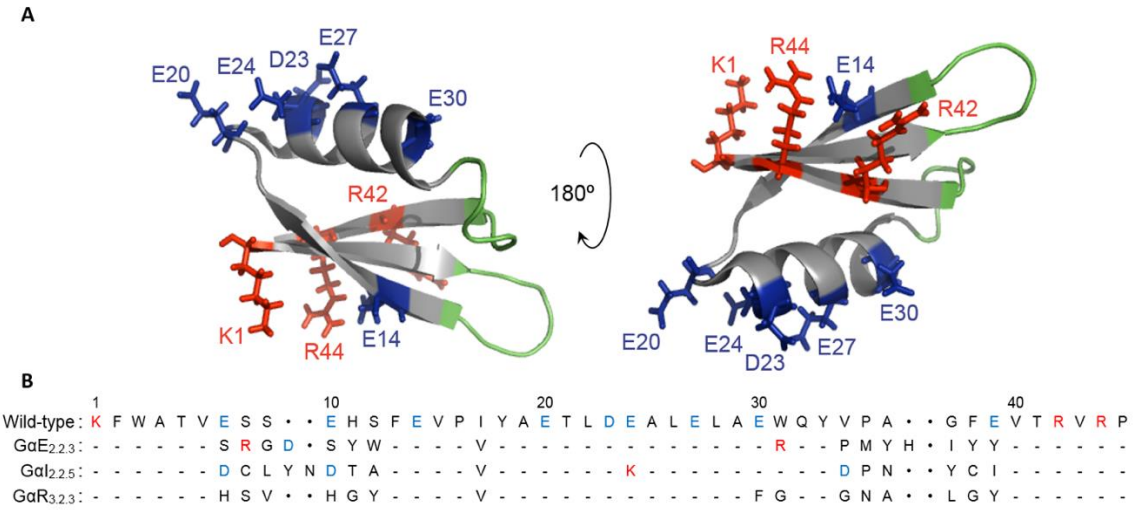



**Figure 4-1.** ***Gp2 charge distribution.*** **(A) Structure of the wild-type Gp2 scaffold with acidic (blue), basic (red), and paratope (green) residues highlighted [PDB: 2WNM[134]] (B) Sequences of wild-type, EGFR-, InsR-, and rIgG-binding clones used as parental ligands for charge modified Gp2 libraries.**

Full amino acid diversity would necessitate $20^{10} = 10^{13}$ unique clones and surpass the transformation efficiency of yeast used for protein display and selection. Therefore, a more selective approach was undertaken to predict which newly introduced mutations would be least likely to impair protein folding and target binding. Three metrics were applied: lack of destabilization as predicted by FoldX[245], frequency in natural homologs[163,164], and amino acid hydrophilicity[246]. FoldX was used to compute the theoretical change in stability upon independent mutation of each charged site (details in *Experimental Procedures*), which predicted that several charged positions would be relatively intolerant to mutation without negatively affecting fold (Figure 4-2A). These include sites 1, 14, 24, 42, and 44 where 9-14 mutations away from wild-type amino acids increased free energy of folding in excess of 0.5 kcal $mol^{-1}$. The remaining charged positions (20, 23, 27, and 30) were more amenable with only 1-6 mutations causing a similar increase in free energy.

Amino acid frequencies from Gp2 homologs (Figure 4-2A), identified via search of available genomes, also provided information regarding potentially acceptable mutations at charged positions. Wild-type residues were generally conserved with an average of 3.2 non-wild-type amino acids observed. This value decreased to 2.4 when excluding replacement with amino acids of identical charge. Sites 1, 14, 24, 30, 42, and 44 had high conservation of wild-type amino acids (≥ 75% consensus) whereas sites 20, 27, and 31 were more amenable to neutralization, accepting several neutral functional groups. Site 23 was highly conserved as a negatively charged residue but relatively indifferent to aspartic (56%) or glutamic (41%) acid.

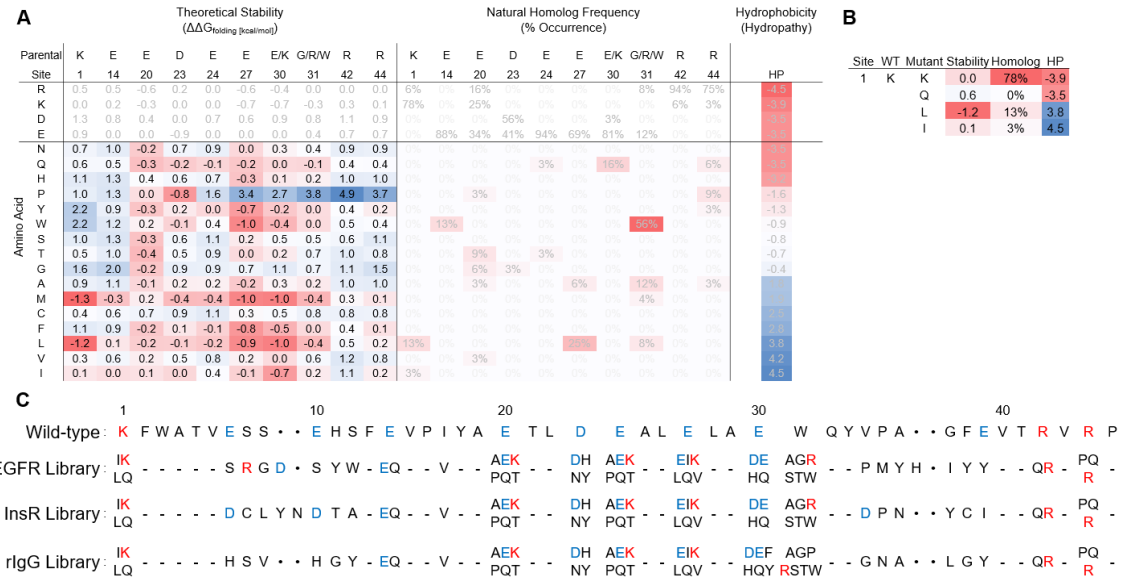



**Figure 4-2. *Combinatorial library design.* (A) Theoretical stability upon mutation based on FoldX[245] force field calculations, natural homolog frequencies from the BLAST algorithm[250], and amino acid hydropathy[246]. (B) Example of the decision-making process behind library designs. (C) EGFR-, InsR-, and rIgG- charge modified Gp2 libraries as designed.**

Each charged position required individual analysis such that no metric (hydrophilicity, stability, and homolog frequency) was applied universally to identify

reasonable mutants for inclusion in the library. All charged functional groups extend outside of the protein core in the wild-type structure[134] making high to moderate hydrophilicity preferable. Additionally, an elevated frequency in naturally occurring homologs guided inclusion of neutral amino acids in the combinatorial library design. Finally, mutations causing a considerable increase in free energy of folding ($\Delta\Delta G_{folding} \geq 1$ kcal $mol^{-1}$), as modeled by FoldX, were avoided unless such mutations presented high homolog frequency (> 5%) and low hydrophobicity (< -1). In some instances, the inclusion of less desirable amino acids was unavoidable due to genetic code restrictions. The following examples are used to demonstrate the overall methodology. Site 1 was allowed to mutate to glutamine for high hydrophilicity, leucine for elevated homolog frequency, and isoleucine as a byproduct of lysine and leucine codons (Figure 4-2B). In some cases, mutations that appeared nondetrimental in one metric would be juxtaposed by the others. For example, methionine was deemed generally acceptable at all sites of interest via stability analysis (median $\Delta\Delta G_{folding,\ mutation}$ = -0.4 kcal $mol^{-1}$) although it is moderately hydrophobic and has a mere 4% homolog frequency at site 31 and 0% frequency elsewhere. Similarly, hydrophilic proline was predicted to be destabilizing at position 44 ($\Delta\Delta G_{folding,mutation}$ = 3.7 kcal $mol^{-1}$), but presented 9% sequence identity in natural homologs. The high throughput of yeast display and flow cytometry for functional selection enabled analysis of many mutants while the combinatorics of diversifying 10 positions required exclusion of many others.

Genes containing the degenerate codons necessary to implement these charged site mutations (Figure 4-2C), unmutated backbone residues, and EGFR-, InsR-, or rabbit IgG-binding paratopes were constructed through overlap PCR of 8-11 oligonucleotides per

library resulting in 0.75 x $10^6$, 0.75 x $10^6$, and 1.7 x $10^6$ possible unique clones. Assembled genes were electroporated into EBY100 *S. cerevisiae* yeast along with a linearized vector to enable yeast display of Gp2 as a C-terminal fusion to Aga2p with a glycine-rich polypeptide linker.[167,255] The resultant yields of 2 x $10^8$, 1.7 x $10^8$, and 2.1 x $10^8$ transformants for EGFR, InsR, and rabbit IgG libraries allow for >99% sampling of theoretical sequence space.[256] Illumina MiSeq deep sequencing of naïve libraries showed that 91%, 94%, and 88% of EGFR, InsR, and rIgG yeast transformants presented sequences as designed. These values increase to 98-99% when single-point mutations are included, indicating high library quality.

### *4.4.2 Mutant Selection*

EGFR, InsR, and rIgG libraries were sorted, via yeast display and flow cytometry, to isolate mutants that retain stability and functional binding. The yeast containing the genetic libraries were grown and induced to express the Aga2p-linker-Gp2 variants tethered to their surface via covalent linkage to the membrane-bound Aga1p mating protein.[167,255] Gp2-displaying yeast libraries underwent consecutive rounds of fluorescence-activated cell sorting (FACS) in which yeast were placed at an elevated temperature to induce unfolding of unstable mutants[177], incubated on ice with target protein, and evaluated for retention of activity. Parental library ligands GαE2.2.3, GαI2.2.5, and GαR3.2.3, each demonstrated target affinities in the low nanomolar range (18 ± 8, 2.4 ± 0.4, and 2.3 ± 1.4 nM) and high thermal stability (71 ± 2 ºC, 78 °C, and 80 ± 2 °C) (Table 4-1).

**Table 4-1. Library sorting conditions for charge mutated Gp2 libraries with parental clone *in vitro* characteristics.**

| Library | Pre-sorting Population | Post-sorting Population | T (°C) | Target (nM) | Parental Clone $T_m$ (°C) | Parental Clone $K_D$ (nM) |
|---|---|---|---|---|---|---|
| **EGFR** | Naïve | #1 | 60 | 20 | 71 ± 2 | 18 ± 8 |
| | #1 | #2a | 60 | 2 | | |
| | #1 | #2b | 80 | 20 | | |
| | #2(a+b) | #3 | 80 | 2 | | |
| | #3 | Final | 4 | deplete non-specific | | |
| **InsR** | Naïve | #1 | 50 | 20 | 78 | 2.4 ± 0.4 |
| | #1 | #2a | 50 | 2 | | |
| | #1 | #2b | 60 | 20 | | |
| | #2(a+b) | Final | 4 | deplete non-specific | | |
| **rIgG** | Naïve | #1 | 50 | 2 | 80 ± 2 | 2.3 ± 1.4 |
| | #1 | #2 | 60 | 2 | | |
| | #2 | Final | 4 | deplete non-specific | | |

These values were used to guide temperature and target concentration stringency during sorts. Naïve libraries exhibited a broad distribution from mutants that lost function to mutants that tolerated thermal stress and exhibited strong target binding (Figure 4-3).

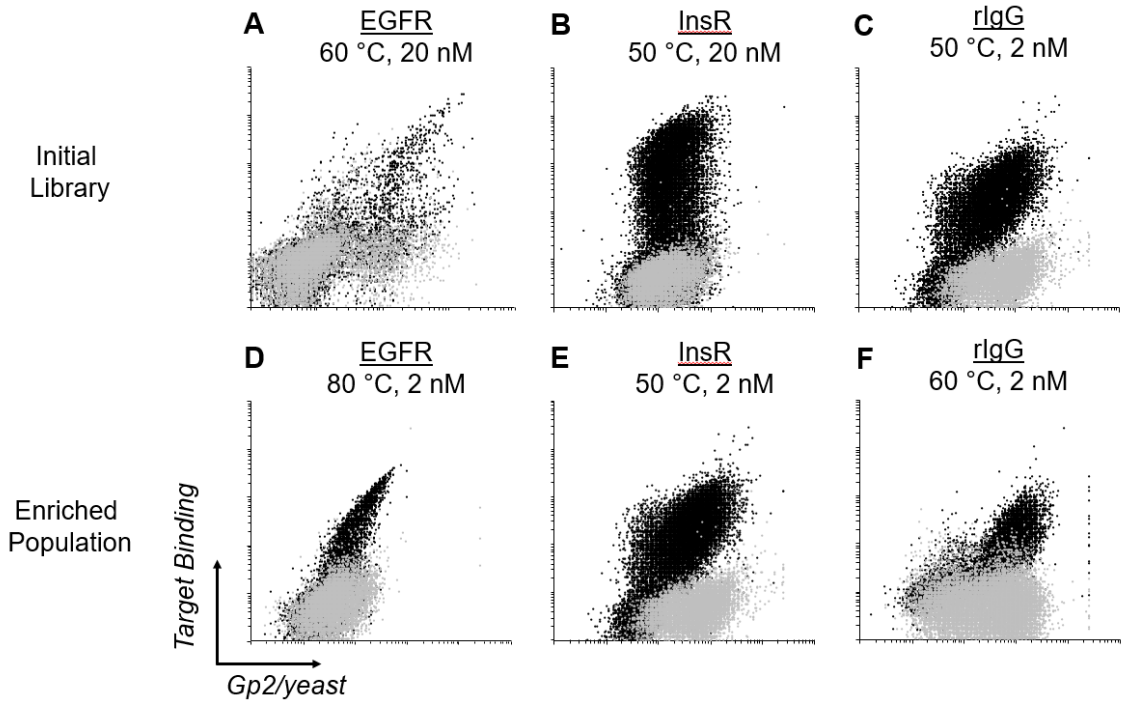


**Figure 4-3.** ***First generation library selections.*** **Naïve (A, B, C) and terminal (D, E, F) GαE, GαI, and GαR populations sorted for stability and binding with (black) and without (gray) their respective recombinant target.**

To remove instances of highly unstable or nonbinding variants, naïve libraries were initially sorted at mild conditions with thermal stress temperatures 10 °C or more below parental $T_m$ and recombinant target at (EGFR, rIgG) or above (InsR) parental dissociation constants (Table 4-1 and *Materials and Methods*). A subsequent set of sorts aimed to isolate populations that either demonstrate improved thermal stability while maintaining moderate affinity (via elevated thermal stress with maintained target concentration) or provide increased ligand affinity after mild thermal stress (maintained thermal stress 11-28 °C below parental $T_m$ while reducing target concentration). To this end, the sorted populations were split and sorted separately at each condition set, then pooled. Compared to GαI and GαR libraries, anti-EGFR clones were quite robust in their response to stability and affinity selection pressures and were sorted an additional time at an elevated temperature (80 °C) and low target concentration (2 nM). Although parental $G\alpha I_{2.2.5}$ and $G\alpha R_{3.2.3}$ exhibit substantial stability (78 and 80 ± 2 °C), no appreciable target-bound populations were present above background signal in sorts exceeding 60 °C even at heightened levels of target. However, unlike GαI, GαR clones retained parental affinity at 60 °C. To deplete clones that exhibit nonspecific binding, enriched populations of Gp2-displaying yeast were incubated with streptavidin- and mouse IgG (mIgG)-coated magnetic beads at saturating conditions[257], and only unbound yeast were retained.

In each of the three rounds of ligand selection against recombinant EGFR, 4-6% of sorted populations were collected based on expression-normalized target binding as indicated by triangular gates (Figure 4-4). To prepare for use in physiological

environments, we aimed to ensure that mutants bound EGFR in its native fold. Thus, the EGFR-binding population underwent two additional rounds of sorting without heating against 2 and 0.2 nM solubilized EGFR from the lysate of A431 epidermoid carcinoma cells (2 – 6% collected).

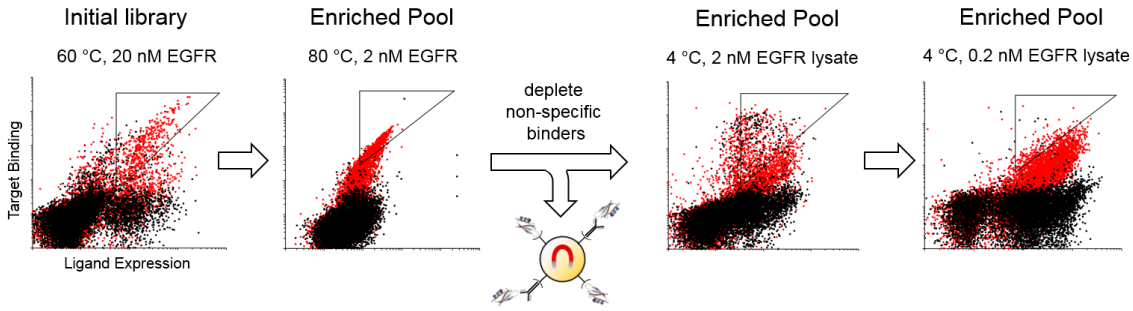



**Figure 4-4.** ***Enrichment of stable, EGFR-binding charge mutants.*** **Yeast libraries displaying charge-modulated Gp2 ligands were sorted against biotinylated EGFR ectodomain (2 – 20 nM) after preheating the yeast for 30 min (60 - 80 °C) followed by screening against streptavidin- and mIgG-coated beads to deplete nonspecific binders, and finally sorted against lysate-extracted EGFR at 2 and 0.2 nM target. EGFR-labeled and target-negative yeast are shown in red and black with collected populations in triangular gates.**

### *4.4.3 Mutant Characterization*

Plasmids from the starting libraries and the concluding functional populations sorted for thermal tolerance and recombinant target binding, after nonspecific Gp2 ligand depletion, were sequenced via Illumina MiSeq. Additionally, 24 clones from solubilized lysate EGFR-sorted populations were sequenced using the Sanger method.[258]

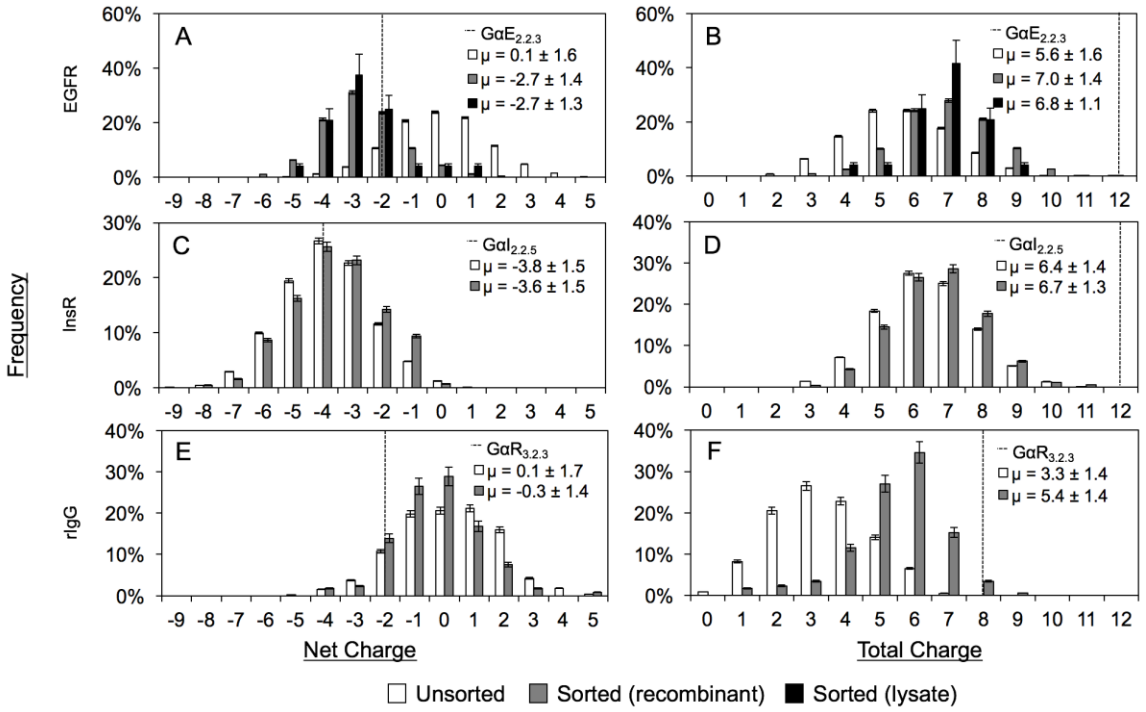



**Figure 4-5.** ***Net and total charge distributions of clones in naïve and sorted populations.*** **(A, B) EGFR (*naïve*: 2705, *sorted*: 3202), (C, D) InsR (*naïve*: 2842, *sorted*: 798), and (E, F) rIgG (*naïve*: 596, *sorted*: 172) Gp2 libraries. Data presented as frequency ± standard deviation. Mean ± standard deviation of overall distribution shown in legend.**

In the EGFR case, negatively charged variants were enriched during functional selection as the mean net charge shifted from neutral (0.1 ± 1.6) in the unsorted library to -2.7 ± 1.4 in the functional pool (Figure 4-5A). Notably, this shift was in the direction of — and past — the parental GαE57, which has a net charge of -2. This net charge shift was achieved via enrichment of wild-type acidic residues as well as neutralization of basic residues: the unsorted library neutralized more than six sites on average (GαE57 has 12 charged sites whereas the unsorted library has 5.6 ± 1.6), which only slightly increased in the functional pool (7.0 ± 1.4), indicating that the larger shift in net charge also had contributions from neutralization of basic residues (Figure 4-5B). Sitewise analysis reveals

the specifics of these functional preferences (Figure 4-6), which are discussed below in the context of all domains.

The other two parental Gp2 domains exhibit divergent tolerances to charge reduction as GαI essentially maintains removed charge (6.7 ± 1.3 charged sites in functional domains relative to 6.4 ± 1.4 in the library and 12 in the parental domain) whereas GαR reverts nearly half of its removed charges (3.3 ± 1.4 naïve to 5.4 ± 1.4 in functional vs. 8 in parental). Akin to the EGFR-targeted domain, but to a lesser extent, GαR net charge evolved toward the negative parental. The lack of change in total charge of GαI is matched by an absence of movement in net charge (-3.8 ± 1.5 in naïve vs. -3.6 ± 1.5 in functional populations). Notably, unlike the other libraries, the GαI library starts as net negative because of four acidic residues and no bases in its evolved loops. One other possible explanation for the tolerance of charge removal in GαI is an increase in structural stability imparted by a predicted disulfide bond between C8 and C38 in the evolved loops.

Across all three targets, the six negatively charged sites exhibit enrichment of the parental acidic side chains, with an increased frequency in functional variants of 29 ± 20%, with E14 and D23 in the InsR-targeted domains being the lone exceptions. Positively charged sites showed greater variance. Positions 1 and 31 display functional preference toward neutral residues frequent in natural homologs. K1 evolved toward leucine in all campaigns, which is present in 13% of native Gp2 homologs. Position 31, which is only charged in the parental EGFR-binding domain, preferentially mutates to W, which is 56% W in native homologs. R42 exhibits nearly comparable preference of Q and R within EGFR-binding domains whereas InsR- and rIgG-binding domains strongly enrich wild-type arginine. R44 exhibits variable performance for each target. EGFR-targeted domains

enrich P and tolerate Q at the expense of R; InsR-targeted domains enrich Q and tolerate R; rIgG-binding domains enrich R and tolerate Q.

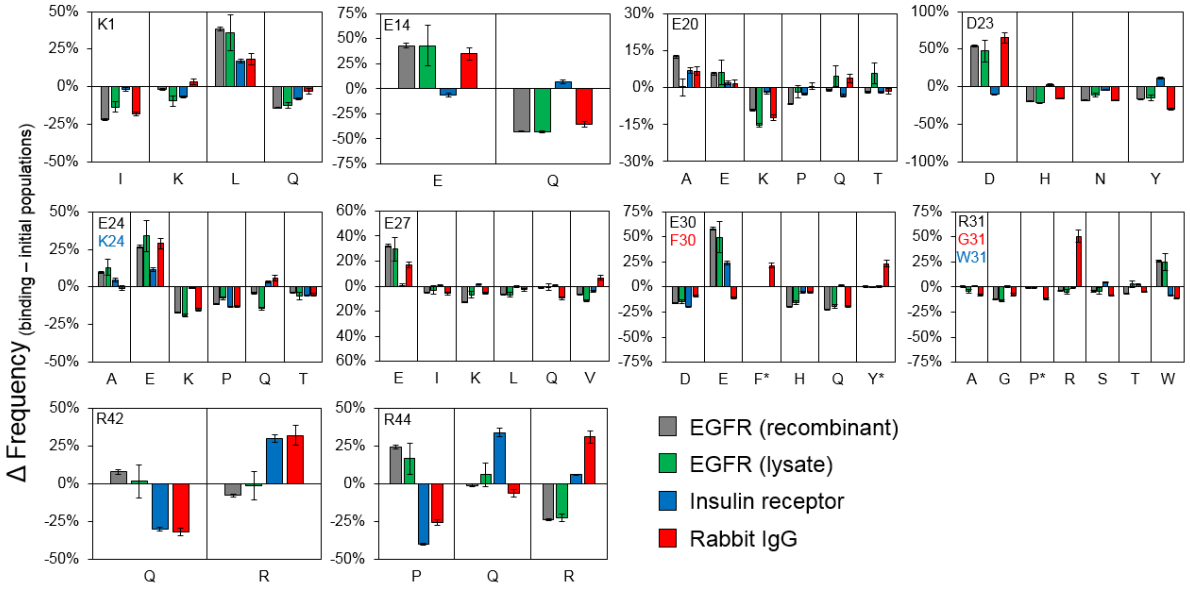



**Figure 4-6. *Sitewise amino acid functionality*. Change in frequencies between functional and naïve populations. Parental amino acids presented in black unless noted by a matching color change in legend (* indicates presence solely in the rIgG library).**

The utility of the combinatorial library approach is evident from the lack of strong correlations between functional efficacy and direct design metrics. Sitewise enrichment of particular amino acids does not correlate with predicted stability by FoldX (Pearson correlation coefficient, R = 0.24) or hydrophilicity (R = 0.01). Consistent with the examples in the previous paragraph, modest correlation is evident for natural homologs (R = 0.47).

Overall, numerous members of each library were robust to thermal denaturation and sorting against concentrations of recombinant target in the low nanomolar range. Additionally, several Gp2 variants bound high picomolar concentrations of A431 lysate-derived EGFR.

### *4.4.4 Top Clones and In Vitro Characterization*

In addition to revealing mutational preferences, these selections identified particular functional mutants. To determine the effects of reducing total charge, six clones, all maintaining parental or near-parental net charge, were chosen from the final EGFR lysate-sorted population to further characterize the impacts of charge modification as well as to identify compelling lead mutants for physiological study.

These variants, with charge states of +2/-2, +2/-4, +2/-5 (two clones), +3/-5, and +3/-6, were produced in T7 *E. coli* and evaluated in vitro for soluble production yield, target affinity, thermal stability, and secondary structure. (Table 4-2).

**Table 4-2. Mutated positions, paratope (loop) sequences, and in vitro characteristics of Gp2 clones recovered in flow cytometry sorts against EGFR derived from A431 cell lysate. $K_d$ values represent equilibrium dissociation constants from titrations of mutants against A431 cells using flow cytometry (n = 3). Midpoint denaturation temperatures ($T_m$) were determined with circular dichroism spectroscopy of purified proteins (n = 1 - 3). Yield is reported as recovered mass of purified protein per L of culture (n = 3).**

| Name | 1 | 14 | 20 | 23 | 24 | 27 | 30 | 31 | 42 | 44 | Loop 1 | Loop 2 | Charge (+/-) | Net Charge | Total Charge | $T_m$ (°C) | $K_D$ (nM) | Yield (mg/L) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GαE22 | L | E | T | Y | A | L | H | S | R | P | SRGDSYW | PMYWHIYY | 2/2 | 0 | 4 | 60 | 6.3 ± 1.4 | 1.8 ± 0.4 |
| GαE24 | L | E | Q | D | T | V | E | W | R | Q | SRGDSYW | PMYWHIYY | 2/4 | -2 | 6 | ND | ND | < 50 |
| GαE25A | L | E | P | D | A | E | E | T | R | P | SRGDSYW | PMYWHIYY | 2/5 | -3 | 7 | 64 | 4.9 ± 0.7 | 1.8 ± 1.2 |
| GαE25B | I | E | E | D | P | Q | E | R | Q | P | SRGDSYW | PMYWHIYY | 2/5 | -3 | 7 | ND | ND | < 50 |
| GαE35 | L | E | Q | D | E | K | E | T | R | P | SRGDSYW | PMYWHIYY | 3/5 | -2 | 8 | 87 ± 1 | 8.8 ± 0.6 | 2.1 ± 0.5 |
| GαE36 | K | E | E | D | E | Q | E | W | R | Q | SRGDSYW | PMYWHIYY | 3/6 | -3 | 9 | 59 | 2.3 ± 1.1 | 5.0 ± 0.3 |
| GαE57 | K | E | E | D | E | E | E | R | R | R | SRGDSYW | PMYWHIYY | 5/7 | -2 | 12 | 79 | 9.4 ± 4.1 | 2.7 ± 0.1 |

Soluble production yield was diminished considerably for GαE24 and GαE25B with less than 0.05 mg/L recoverable for either clone. However, recovery in the remaining

four mutants exceeded 1 mg/L allowing for further in vitro characterization. As selected for, all four mutants exhibited good binding affinity (Figure 4-7).

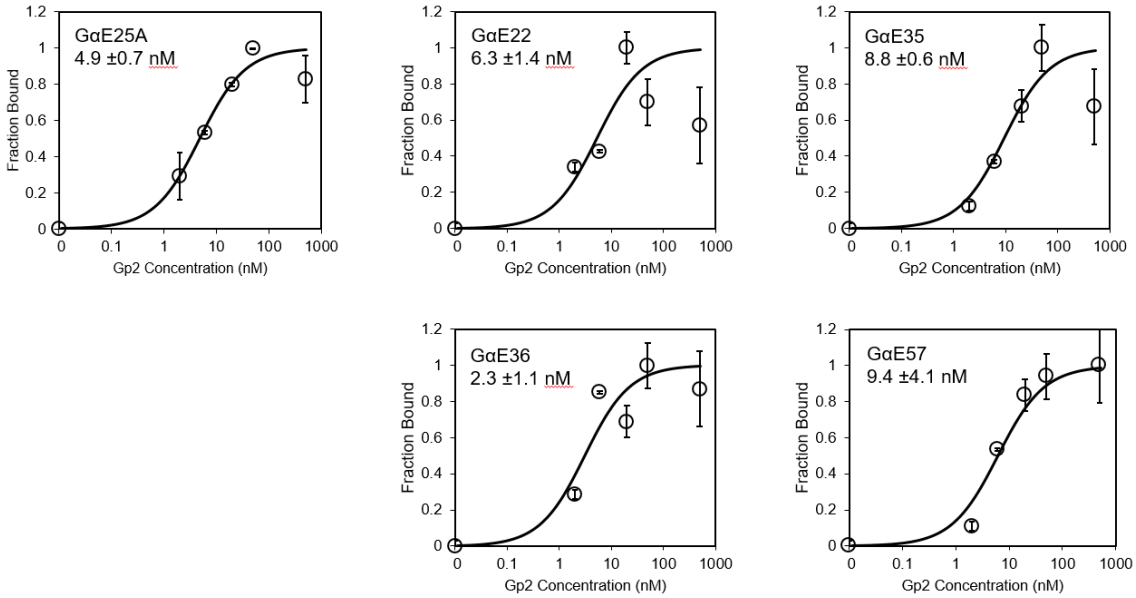



**Figure 4-7. *Anti-EGFR Gp2 affinity titrations.* A431 cells were incubated on ice with 2, 6, 20, 50, or 500 nM Gp2 ligand followed by labeling with fluorescein-conjugated anti-His6 antibody and median fluorescence analysis via flow cytometry. Equilibrium dissociation constants were calculated using a 1:1 binding model (n = 3).**

Compared to GαE57, charge reduced clones GαE25A and GαE36 demonstrated nominally improved target affinity against recombinant EGFR (9.4 ± 4.1 nM for parental vs. 4.9 ± 0.7 nM (GαE25A) and 2.3 ± 1.1 nM (GαE36)) while GαE22 and GαE35 maintained near parental values (6.3 ± 1.4 nM and 8.8 ± 0.6 nM). Conversely, the mutants diverged with regard to thermal stability. GαE35 improved upon parental stability ($T_m$ = 87 ± 1 °C) whereas GαE25A, GαE22, and GαE36 saw 15-20 °C reductions in melting temperature (Figure 4-8).

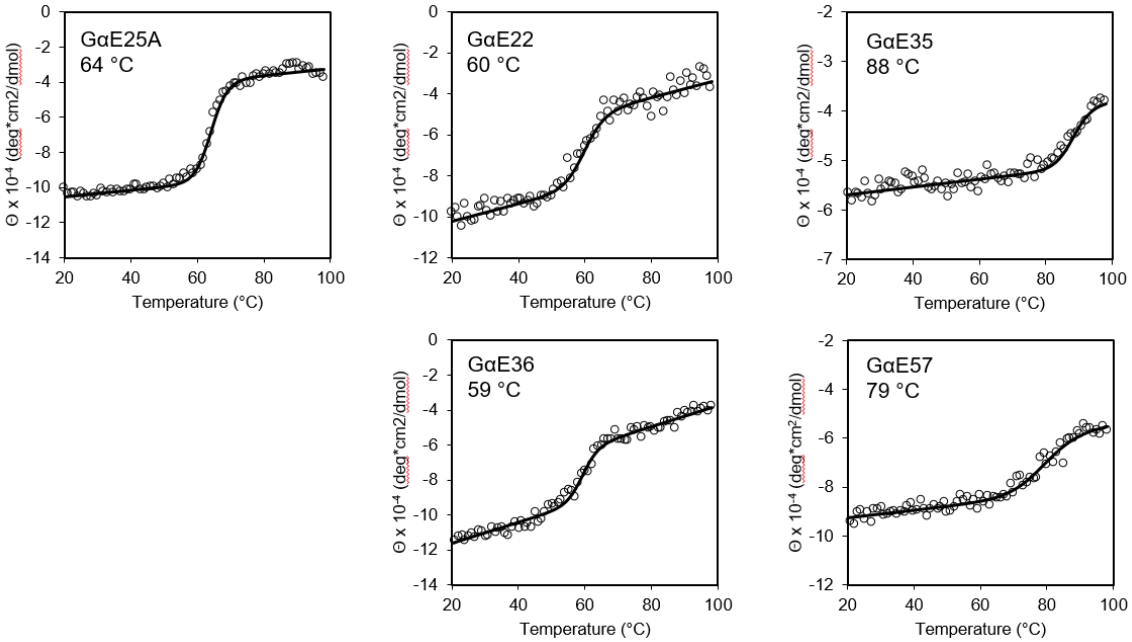



**Figure 4-8.** ***Anti-EGFR Gp2 thermal denaturation curves.*** **Purified ligands were scanned at a wavelength of 220 nm during heating from 20 to 98 °C (2 °C/min). The midpoint of denaturation ($T_m$) was determined using a two-state protein unfolding model.**

Secondary structure was reasonably maintained in all variants (Figure 4-9). GαE25A showed a slight shift toward increased α-helical content. Interestingly, clones that did not fully return to their original secondary structure after heating (GαE25A, GαE22, GαE36) also exhibited a demonstrable decrease in $T_m$ when compared to fully refoldable clones (mean ± SD: 61 ± 3 °C vs 83 ± 6 °C).

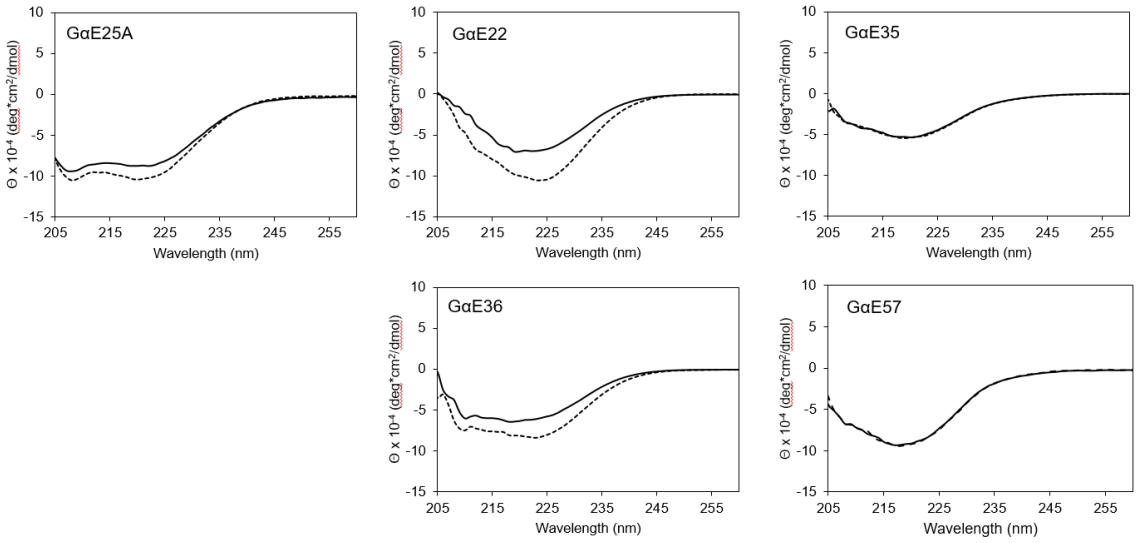



**Figure 4-9.** ***Circular dichroism spectroscopy of recovered mutants.*** **Purified anti-EGFR Gp2 mutants were analyzed by circular dichroism spectroscopy in triplicate between 205 and 260 nm wavelengths before (- - -) and after (——) thermal denaturation and cooling.**

Five of the six clones analyzed were either not appreciably producible or saw a significant reduction in $T_m$. This suggests that more stringent thermal denaturation temperatures prior to sorting against target and perhaps an increase in the number of individually assessed clones may be required to isolate more useful ligands. Nevertheless, a mutant (GαE35) was identified that reduces total charge by four and alters charge distribution (Figure 4-10) while maintaining net charge, elevates thermal stability, maintains binding affinity, and retains recombinant yield. In addition, a mutant (GαE22) was identified with greatly reduced total charge and maintained binding affinity albeit with notably lower thermal stability.

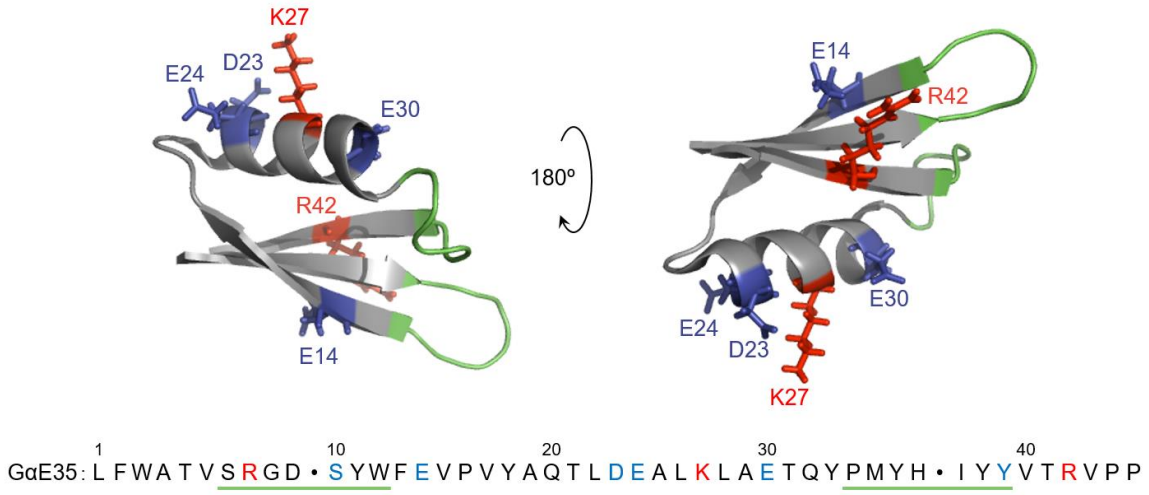



**Figure 4-10.** ***Expected structure and sequence of GαE35.*** (A) Approximate structure, based on PDB: 2WNM[134], and sequence of the GαE35 ligand with acidic (blue), basic (red), and paratope (green) residues highlighted.

### *4.4.5 Physiological Distribution and Protease Degradation*

To ascertain the effects that reduced and redistributed charge may impart on an EGFR-binding ligand in vivo, GαE22, GαE35, and parental GαE57 were chosen as candidates for imaging of tumor-bearing murine models. GαE35 retains parental net charge of -2, providing an opportunity to elucidate the effects of a reduction in total charge. GαE22 is the least charged anti-EGFR Gp2 isolated from the selections creating a more extreme comparison between GαE57 and a substantially charge-reduced variant. To address target specificity, A431 ($(2.9 \pm 1.4) \times 10^6$ EGFR per cell[227]) and MDA-MB-435 ($(1.5 \pm 1.1) \times 10^4$ EGFR per cell[227]) were incubated with 500 nM GαE22, GαE35, or GαE57 and analyzed via flow cytometry. Binding to A431 cells was substantially higher than to MDA-MB-435 cells for GαE22, GαE35, and GαE57 ($p \leq 0.001$) (Figure 4-11).

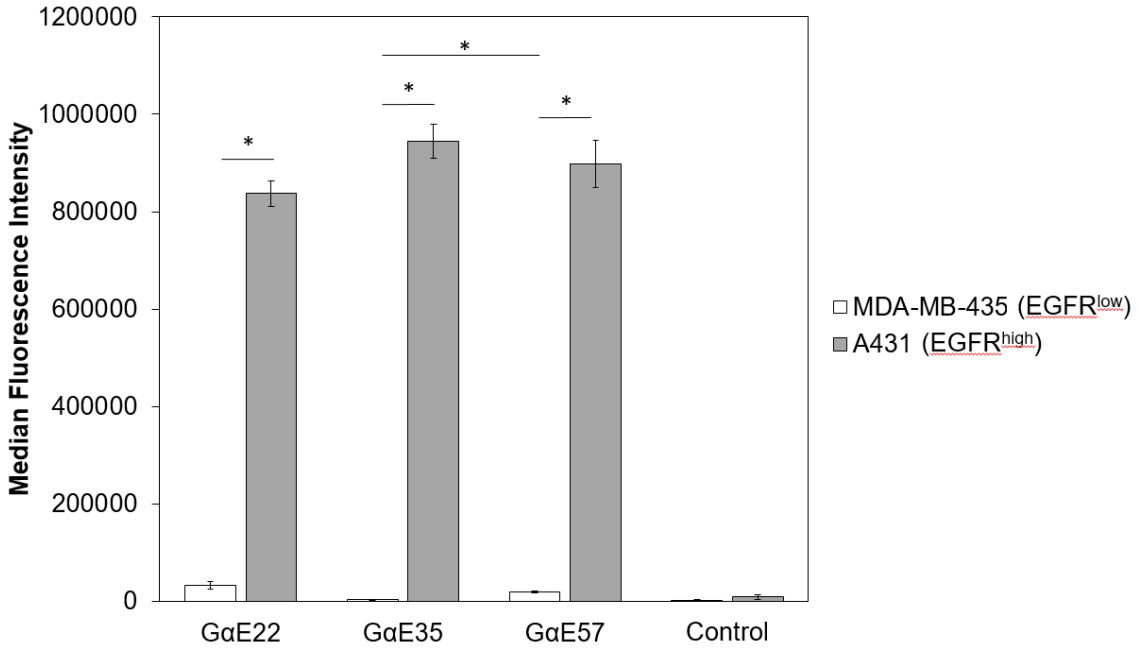



**Figure 4-11.** ***EGFR binding specificity.*** **MDA-MB-435 ($EGFR^{low}$) and A431 ($EGFR^{high}$) cells were incubated on ice with 500 nM Gp2 ligand, labeled with fluorescein-conjugated anti-$His_6$ antibody, and median fluorescence signal determined with flow cytometry (n = 3) (* signifies $p \leq 0.001$).**

While all clones exhibited strong differentiation between the cell types, binding to $EGFR^{low}$ MDA-MB-435 cells was significantly lower for GαE35 than the parental GαE57 ($p < 0.001$), and at a ratio consistent with reduced EGFR expression, which suggests enhanced specificity for the charge-engineered GαE35. These clones were conjugated with the metal ion chelator NODAGA through an amine-reactive cross-linking reaction with its N-hydroxysuccinimidyl ester. Labeling was nonspecific due to the free amines present on lysines at position 1 of GαE57 and 27 of GαE35. However, conjugation was on average singular with $0.9 \pm 0.2$, $1.1 \pm 0.2$, and $1.1 \pm 0.4$ NODAGA per molecule for GαE22, GαE35, and GαE57 (Figure 4-12).

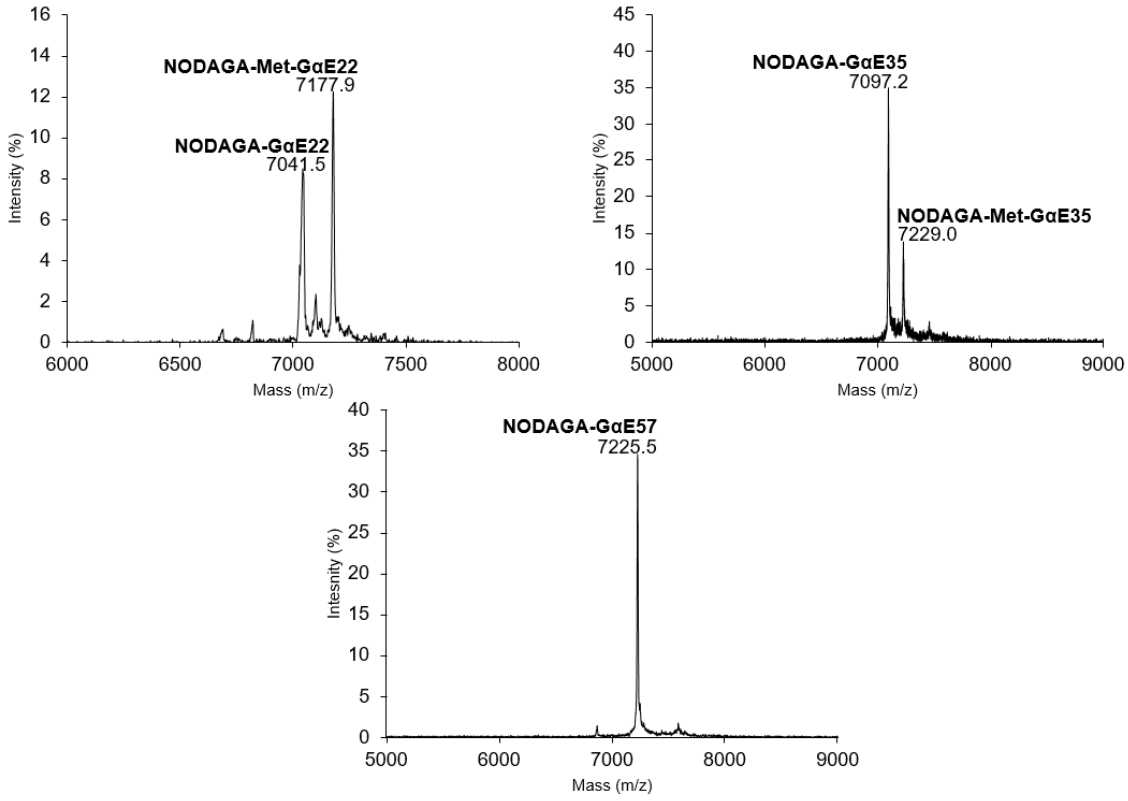



**Figure 4-12.** ***Conjugation of NODAGA to Gp2 charge variants.*** **Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) of NODAGA-conjugated GαE22, GαE23, and GαE57.**

NODAGA-conjugated ligands were mixed with 1-2 mCi of $^{64}$Cu in 100 mM sodium acetate, pH 6.0 at 42 °C for 60 min. Complexed $^{64}$Cu was separated from free copper via size exclusion filtration resulting in 38-48% labeling efficiency and 87-98% radiochemical purity (Figure 4-13).

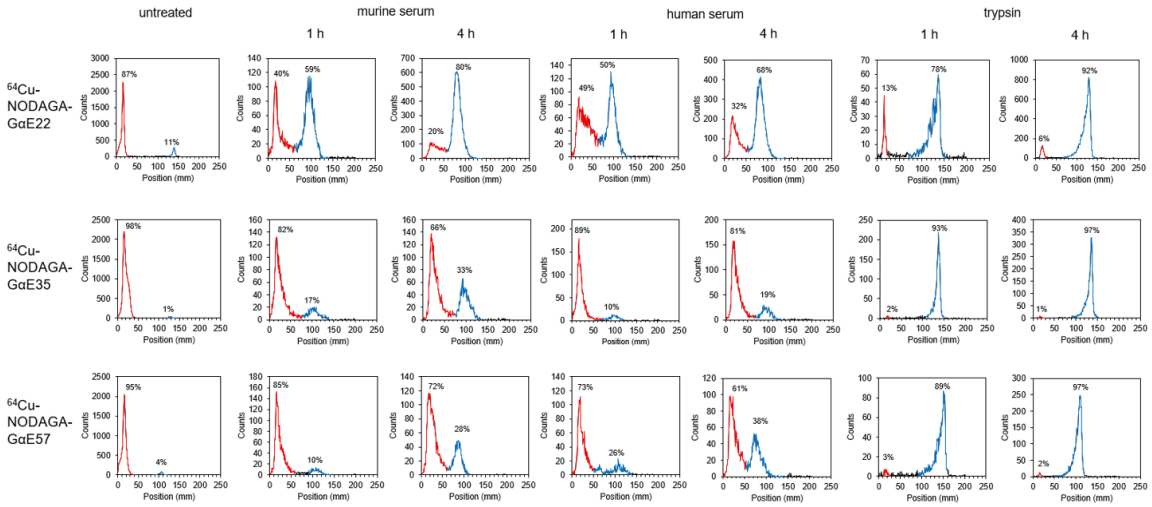



**Figure 4-13. *Proteolytic stability*. Radio-thin-layer chromatography of radiolabeled GαE clones left untreated, in serum, or trypsin.**

Based on yields from previous purifications of unlabeled Gp2 ligands, approximate specific activities were calculated as 0.5-1.2 MBq nmol-1. Physiological biodistribution of each radiolabeled ligand was ascertained by injecting 1.3 ± 0.3 MBq into mice (n = 5-6) bearing $EGFR^{high}$ A431 and $EGFR^{low}$ MDA-MB-435 tumors. Two hours after injection tissues were excised, weighed, and measured for radioactivity (Figure 4-14). $^{64}$Cu-NODAGA-GαE57 and $^{64}$Cu-NODAGA-GαE35 localized strongly to EGFR-overexpressing A431 cells, displaying 5.5 ± 0.5 %ID/g and 4.7 ± 0.5 %ID/g. Binding was target specific ($p < 0.01$) with 0.9 ± 0.1 %ID/g and 0.6 ± 0.1 %ID/g on $EGFR^{low}$ MDA-MB-435 tumors for $^{64}$Cu-NODAGA-GαE57 and $^{64}$Cu-NODAGA-GαE35, respectively. This improved in vivo specificity of $^{64}$Cu-NODAGA-GαE35 ($p = 0.01$) is consistent with reduced nonspecific binding in vitro (Figure 4-11). $Tumor^{EGFR\text{-}high}$:$tumor^{EGFR\text{-}low}$, tumor:blood, and tumor:muscle ratios were all strong and nominally higher for the charge-reduced $^{64}$Cu-NODAGA-GαE35 (Figure 4-14, inset). As with many small protein probes, renal signal is elevated. $^{64}$Cu-NODAGA-GαE57 has modestly lower renal signal than $^{64}$Cu-

NODAGA-GαE35, 153 ± 25 %ID/g vs. 209 ± 18 %ID/g, respectively. Liver activity is slightly elevated for both $^{64}$Cu-NODAGA-GαE57 (2.6 ± 0.1 %ID/g) and -GαE35 (3.0 ± 0.3 %ID/g), but background signal remains below 2 %ID/g for the remaining tissues sampled.

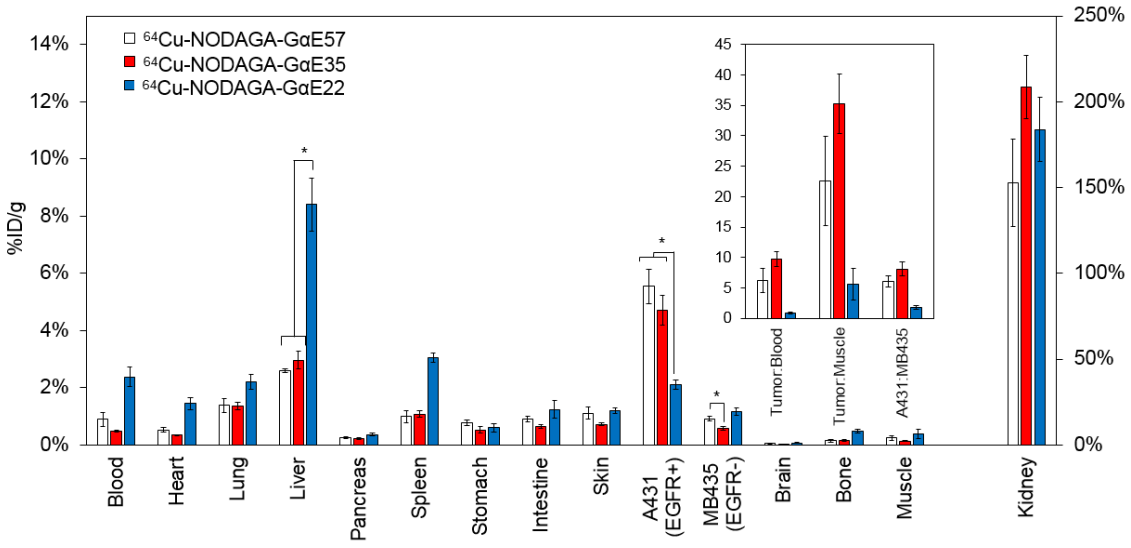



**Figure 4-14.** ***Excised tissue analysis of charge modified GαE ligands.*** **Mice xenografted with A431 (EGFR$^{high}$) and MDA-MB-435 (EGFR$^{low}$) tumors on opposing shoulders were intravenously injected with $^{64}$Cu-NODAGA-GαE57, $^{64}$Cu-NODAGA-GαE35, or $^{64}$Cu-NODAGA-GαE22 and euthanized 2 h post-injection. Tissues or fluid of interest were extracted, weighed, and gamma decay radioactivity measured (%ID/g) (n = 5 – 6). *: p < 0.05. Inset represents ratios of tumor:background signal.**

Conversely, $^{64}$Cu-NODAGA-GαE22 was not an effective EGFR ligand *in vivo*. A431 tumor signal was nearly 2-fold lower than that of GαE57 and GαE35 clones. Additionally, off-target retention was statistically higher in blood, heart, lung, liver, spleen, and bone samples ($p < 0.05$). Background signal was most pronounced in the liver where $^{64}$Cu-NODAGA-GαE22 displayed 8.4 ± 1.0 %ID/g, approximately 3-fold higher than $^{64}$Cu-NODAGA-GαE57 and $^{64}$Cu-NODAGA-GαE35. While this broad background could be related to slightly increased nonspecific binding (Figure 4-11), the co-occurrence of reduced EGFR$^{high}$ tumor localization is more consistent with molecular instability. The

propensity to enzymatically degrade was determined *in vitro* by adding $^{64}$Cu-NODAGA-GαE22, -GαE35, and –GαE57 to murine serum, human serum, or 0.25% EDTA-trypsin at 1:20 volumetric ratios and heating at 37 °C for 4 h. Radio-thin-layer chromatography of samples taken at 1 and 4 hours post-mixing was used to determine the relative amounts of proteolytic metabolites (Figure 4-13). $^{64}$Cu-NODAGA-GαE22 was highly unstable with 59% and 50% cleavage in human and murine serum after 1 h. These values increased to 80% and 68% by 4 h. $^{64}$Cu-NODAGA-GαE35 and $^{64}$Cu-NODAGA-GαE57 were more resistant to enzymatic degradation in serum. Proteolysis of $^{64}$Cu-NODAGA-GαE35 and $^{64}$Cu-NODAGA-GαE57 in human and murine serum resulted in 10% and 17% and 10% and 26% degradation by 1 h, respectively. All clones were degraded to near completion (≥ 92%) in EDTA-trypsin after 4 h. In retrospect, the proteolytic instability of GαE22 may have been predicted by its decreased $T_m$ compared to GαE35 and GαE57.

### *4.4.6 Positron Emission Tomography Imaging of GαE35*

In contrast to parental $^{64}$Cu-NODAGA-GαE57, $^{64}$Cu-NODAGA-GαE35 showed statistically lower MDA-MB-435 signal, similar A431 retention and off-target uptake, and nominal improvements in tumor:blood and tumor:muscle ratios. Buoyed by these results, PET/CT imaging of $^{64}$Cu-NODAGA-GαE35 was performed in murine models bearing A431 xenografts, with dual tumors to increase experimental observations of tumor uptake and retention. To further evaluate target specificity, imaging was performed with and without preblocking by an injection of unlabeled NODAGA-GαE35. Copper chelation of $^{64}$Cu-NODAGA-GαE35 and purification were performed as previously described resulting in 66% labeling efficiency, 97% radiochemical purity, and 1.43 ± 0.01 MBq nmol$^{-1}$ specific

activity. To demonstrate native EGFR specificity, four mice were intravenously injected with unlabeled NODAGA-GαE35 at 87 – 144-fold molar excess (0.7 – 0.8 mg) followed by 1.5 ± 0.4 MBq of $^{64}$Cu-NODAGA-GαE35 between 5 and 10 min later. Four separate mice were injected with 1.9 ± 0.7 MBq of $^{64}$Cu-NODAGA-GαE35 without blocking by unlabeled ligand. CT/PET scans were performed at 1 and 2 h post-injection (Figure 4-15A). Using CT scans as an anatomical guide liver, muscle, tumor, and kidney signals were quantified through PET image analysis at both time points (Figure 4-15B). $^{64}$Cu-NODAGA-GαE35 localized to EGFR-overexpressing A431 tumors (5.6 ± 0.6 %ID/g) and cleared background (7.2 ± 1.1 tumor:muscle ratio) within 60 min. Tumor signal (6.2 ± 0.6 %ID/g) and tumor:muscle ratio (6.7 ± 1.1) were essentially steady at 2 h p.i.

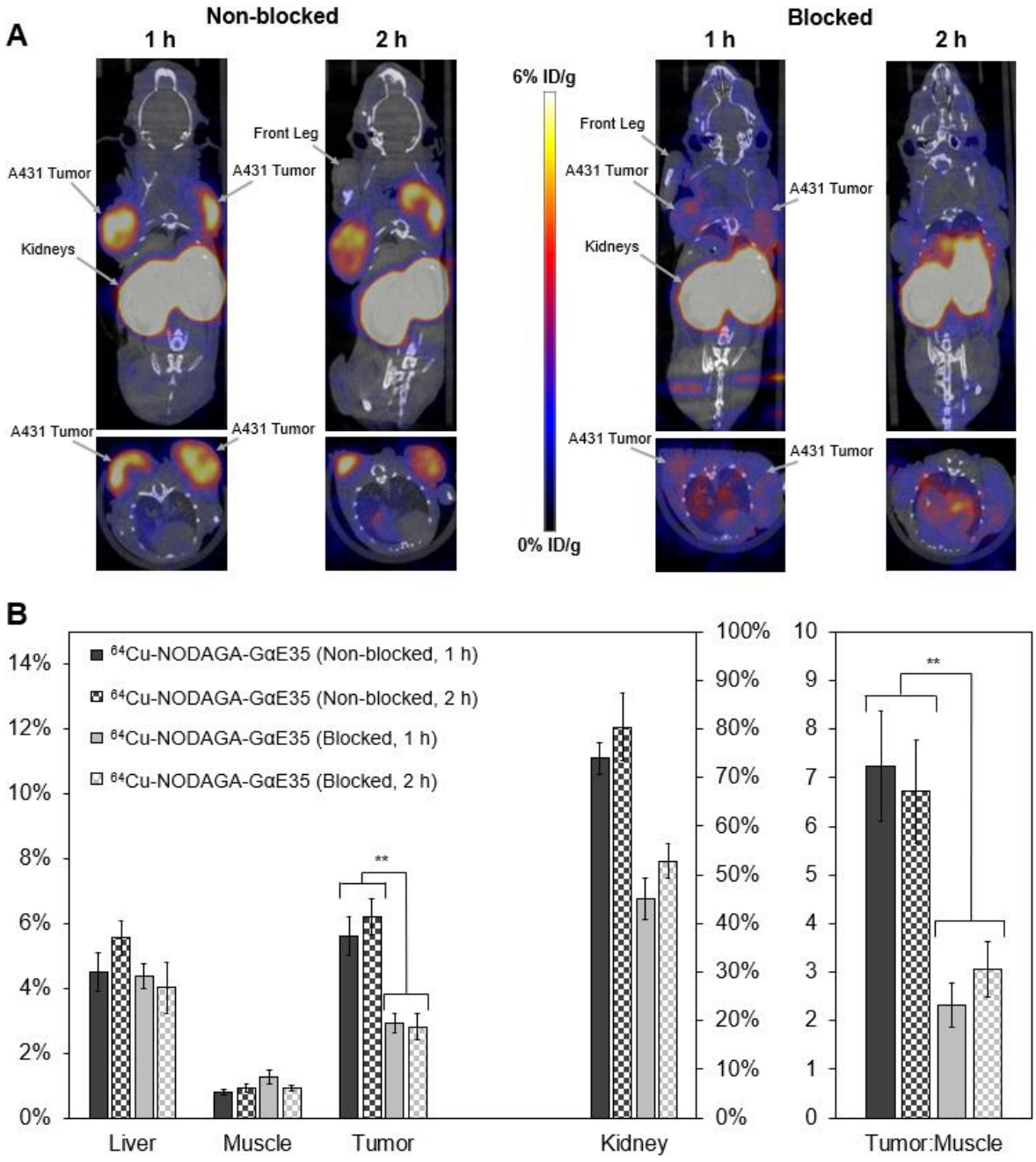



**Figure 4-15.** ***PET/CT imaging and PET image quantification.*** **(A) Mice xenografted with two A431 (EGFR$^{high}$) tumors on opposing shoulders were intravenously injected with either $^{64}$Cu-NODAGA-GαE35 (Nonblocked) or NODAGA-GαE35 at (102 ± 29)-fold molar excess followed by $^{64}$Cu-NODAGA-GαE35 5 - 10 min later (Blocked) and imaged via PET/CT at 1 and 2 h post-injection. Coronal and transverse slices through tumors are shown for representative mice (n = 4). (B) Signal in the liver, muscle, tumors, and kidney were quantified using Siemens Inveon research software at 1 and 2 h post-injection (**: $p < 0.01$).**

Binding was target specific as demonstrated by a nearly 2-fold reduction in tumor signal when pre-blocked with excess NODAGA-GαE35 at both 1 (2.9 ± 0.3 %ID/g) and 2 h (2.8 ± 0.4 %ID/g) ($p < 0.01$). Muscle retention was low and indifferent to injection scheme at 1 h (1.3 ± 0.2 vs. 0.8 ± 0.1 %ID/g) and 2 h (0.9 ± 0.1 vs. 0.9 ± 0.1 %ID/g) for mice with or without cold NODAGA-GαE35 blocking. Hepatic signal was elevated relative to muscle in blocked (4.4 ± 0.4 %ID/g) and nonblocked (4.5 ± 0.6 %ID/g) models after 1 h, remained constant at 2 h in blocked models (4.0 ± 0.8 %ID/g), and increased modestly at 2 h in nonblocked mice (5.6 ± 0.5 %ID/g). Consistent with biodistribution analysis of other small protein ligands[94,129,214,254,259], kidney signal is elevated due to renal processing. Renal uptake was 1.5x and 1.6x higher in nonblocked mice at 1 and 2 h. Note that due to high levels of photon emission within the kidneys, multiple coincident events may be observed within the given collection time window and therefore discarded, leading to an artificially lowered signal.

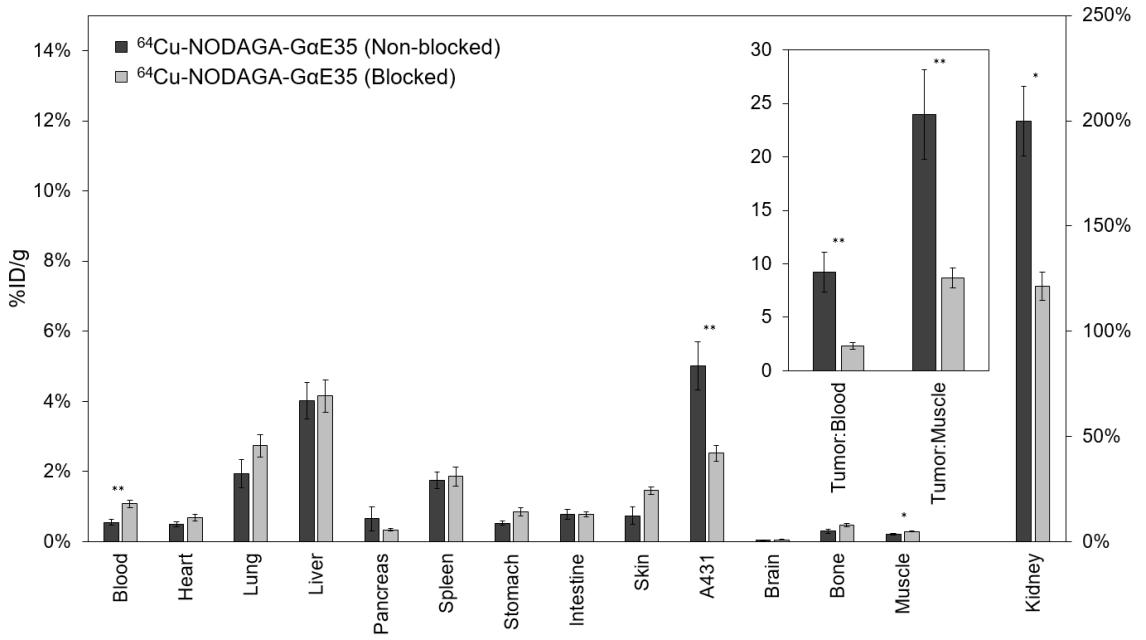



**Figure 4-16.** ***Excised tissue analysis of $^{64}$Cu-NODAGA-GαE35 with and without blocking by NODAGA-GαE35.*** **Mice (n = 4) xenografted with two A431 (EGFR$^{high}$) tumors on opposing shoulders were intravenously injected either $^{64}$Cu-NODAGA-GαE35 (Nonblocked) or NODAGA-GαE35 at (102 ± 29)-fold molar excess followed by $^{64}$Cu-NODAGA-GαE35 5 - 10 min later (Blocked). Mice were euthanized 2 h post-injection. Tissues or fluid of interest were extracted, weighed, and gamma decay radioactivity measured (%ID/g) (*: p < 0.05; **: p < 0.001).**

Two hours post-injection, mice were euthanized with tissues of interest excised, weighed, and measured for radioactivity (Figure 4-16). PET image quantifications were corroborated by tissue measurements. $^{64}$Cu-NODAGA-GαE35 exhibited strong EGFR-specific uptake (5.0 ± 0.7 %ID/g nonblocked vs. 2.5 ± 0.2 %ID/g blocked; $p < 0.001$), elevated tumor:blood (9.2 ± 1.9 %ID/g nonblocked vs. 2.3 ± 0.3 %ID/g blocked; $p < 0.001$) and tumor:muscle (25 ± 11 %ID/g nonblocked vs. 8.7 ± 2.4 %ID/g blocked; $p < 0.001$) ratios. $^{64}$Cu-NODAGA-GαE35 tumor signal was nominally higher when quantified via image analysis versus excised tissue scintillation (6.2 ± 0.6 %ID/g vs. 5.0 ± 0.7 %ID/g), although this difference was not statistically significant. Additionally, muscle signal was

slightly increased in both blocked (0.9 ± 0.1 %ID/g vs. 0.3 ± 0.0 %ID/g) and nonblocked (0.9 ± 0.1 %ID/g vs. 0.2 ± 0.0 %ID/g) images.

## 4.5. **Conclusions**

We created combinatorial libraries through rational design to introduce charge-neutralizing or charge-flipping mutations on the Gp2 scaffold. By briefly heating yeast-displayed Gp2 domains to denature unstable clones followed by flow cytometry to sort for the retention of binding, deep sequencing of the enriched populations revealed preferred mutations across three diverse binders (to EGFR, InsR, and rIgG). Functional populations indicated tolerance of substantial charge removal with a propensity toward reintroduction of charge and a maintenance of negative net charge. Additional screening identified a promising EGFR Gp2 ligand with improved charge distribution, reduced charge density, high stability, and strong, specific binding. Three lead EGFR binders, GαE22, GαE35, and parental GαE57, were conjugated with NODAGA, complexed with $^{64}$Cu, and analyzed *in vivo* using murine models bearing EGFR$^{high}$ and EGFR$^{low}$ tumors. NODAGA-GαE35 demonstrated good EGFR$^{high}$ tumor localization (4.7 ± 0.5 %ID/g) and, relative to parental NODAGA-GαE57, improved specificity with reduced EGFR$^{low}$ uptake (0.6 ± 0.1 vs. 0.9 ± 0.1 %ID/g, $p = 0.01$) and nominally improved tumor:blood (9.7 ± 1.3 vs. 6.2 ± 2.0), tumor:muscle (35 ± 5 vs. 23 ± 7), and EGFR$^{high}$:EGFR$^{low}$ (8.1 ± 1.2 vs. 6.1 ± 0.9) ratios. Further study with additional mutants will be needed to more broadly assess the impact of charge reduction and redistribution on ligand performance. Our study allows for the enhancement of the Gp2 framework through guided charge modulation, validates a method

to engineer these modifications, and provides a lead clone, GαE35, to advance physiological EGFR targeting for the purpose of molecular imaging.

### 4.6. Acknowledgements

This work was supported by grants from the National Institutes of Health (R21 EB021511 and R01 EB023339 to B.J.H.) and a University of Minnesota Doctoral Dissertation Fellowship (to M.A.K.). We would also like to thank Dr. Jacob Petersburg (College of Pharmacy, University of Minnesota) for assisting in murine tail vein injections, Patrick Holec for computing FoldX stabilities, and Dr. Tim Starr (University of Minnesota) and Dr. Daniel Vallera (University of Minnesota) for providing MDA-MB-435 melanocyte and A431 epidermoid carcinoma cell lines.

# Chapter 5 - Concluding Remarks and Future Work

Protein charge – net, density, and distribution – is an important physicochemical property that impacts developability and efficacy of biologics. Yet charge manipulation is a challenging task, particularly in the context of small protein ligands, which otherwise provide advantages in physiological distribution relative to antibodies, particularly for diagnostic molecular imaging. This thesis describes the charge engineering of affibody and Gp2 scaffolds through the creation of rationally guided clones and combinatorial libraries which were screened for stability, recombinant yield, and target affinity; deep-sequenced to determine preferred charge distributions and site-specific mutations; and lead clones evaluated *in vivo* as molecular imaging probes for the commonly overexpressed cancer biomarker, EGFR.

To evaluate the impact of protein charge on the affibody scaffold, a hybrid clone 'EA68' was formed through the incorporation of an anti-EGFR paratope into a stabilized affibody framework. Of the charge engineering techniques utilized to modify EA68, synthetic consensus design proved most robust, providing a clone (EA35S), which neutralized three acidic and three basic residues, with yields and stability comparable to EA68 and an increased EGFR affinity (1.7 ± 0.5 nM vs. 5.3 ± 1.7 nM). The utility of synthetic consensus design was further validated through its application in the creation of additional EA35S-based clones carrying +6/-2 and +2/-6 charge states which retained parental EGFR affinity and secondary structure with only a slight reduction in thermal stability for the +6/-2 variant. Charge engineering of the Gp2 scaffold relied upon diverse combinatorial libraries for charge-neutralizing and charge-flipping mutations introduced

into InsR-, rIgG-, and EGFR-targeted domains. Enriched populations demonstrated target-dependent total, net, and site-specific charge preferences. Average formal and total charge states in recovered populations typically migrated towards or remained at parental values, but differences across libraries suggest charge-modifying diversity may not be uniformly accepted even within the same scaffold type. These results reinforce the utility of synthetic consensus design as a puissant tool for protein charge engineering with neutral or beneficial impact on primary function and biophysical properties. Additionally, site-specific amino acid regulation across libraries with varied targets may be a better predictor of globally accepted mutations, providing insight into an improved scaffold framework for novel clones against additional biomarkers of interest.

During independent affibody and Gp2 experiments, lead anti-EGFR ligands EA35S, GαE35, and their charge-varied homologs were evaluated as PET imaging agents in mouse models bearing $EGFR^{high}$ and $EGFR^{low}$ xenografts to determine the effect of total, net, and distribution of charge on ligand efficacy *in vivo*. A comparison of radiometal chelator moieties using EA35S and EA68 affibodies revealed that, when compared to DOTA, $^{64}Cu$ complexation with NODAGA yielded a simultaneous increase in tumor retention and decrease in background tissue signal, elevating disease state imaging contrast. The sole net positive affibody tested, EA62S, exhibited 2 – 3-fold higher liver and spleen signal comparatively. In contrast to GαE35 and GαE57 – and ostensibly due to molecular instability – net neutral GαE22 performed poorly and presented less than half of the tumor signal with significantly higher background tissue retention. Although further experimental validation in required, these outcomes advocate for the monitoring of proteolytic susceptibility in charge mutants and indicate that an increased positive charge may be lead

to undesirable hepatic retention. The greatest overall *in vivo* performance was achieved by $^{64}$Cu-NODAGA-EA35S and $^{64}$Cu-NODAGA-GαE35 at 4 and 2 h p.i., respectively. EA35S and GαE35 displayed high (12 ± 2 %ID/g and 5.5 ± 0.5%ID/g) and specific (tumor:muscle = 34 ± 5 and 35 ± 5; tumor:blood = 12 ± 3 and 9.7 ± 1.3) EGFR$^{high}$ tumor localization.

EA35S and GαE35 show promise as molecular imaging agents and their clinical implementation may address the need for noninvasive, patient stratifying PET probes. However, dose toxicity and immunogenicity should be determined to prevent adverse patient side effects. A direct *in vivo* comparison between EA35S, GαE35, and an alternative EGFR-contrast agent such as Ac-Cys-$Z_{EGFR:1907}$[213] which is currently under clinical evaluation as a PET/CT radiotracer for non-small-cell lung cancer patients would elucidate the probe most likely to succeed clinically. Given its proven stability and efficacy in murine models the NODAGA chelator should be retained, but experiments should transition from $^{64}$Cu to $^{68}$Ga as radioisotope. Benefits of gallium usage include facile, on-site production from $^{68}$Ge/$^{68}$Ga generators[260] and a comparatively shorter half-life (~1.1 vs 12.7 h), making it suitable for PET imaging at early time points (< 2 h) while decreasing the potential for day-scale radioactive dosing to off-target tissues.

As evidenced by differences in net, total, and site-specific ionic charge biases between three, target-distinct Gp2 libraries (Chapter 4), the distribution of framework charge may be paratope specific and indicative of a need to sample broader charge-mutating sequence space in order to recover ligands with both target affinity function and desired physicochemical properties. Sequence consensus between populations screened against disparate targets illuminates globally acceptable charge modifications while individually upregulated clones lead to affinity ligands for molecular recognition of disease

state biomarkers. Ongoing and future work combines evolved affibody[261] and Gp2[262] paratope regions with scaffold frameworks containing desired redistributions and reductions of ionizable residues based on previously determined synthetic consensus frequencies.

Ionic charge affects numerous protein properties yet it is rarely considered during discovery and ex post facto modifications are often detrimental to function. The field of protein engineering would benefit from preliminary consideration of charge states during initial phases of protein design and discovery. Domains with high charge densities and structures containing isolated cationic and anionic regions should be avoided. However, this type of prescreening may not always be feasible for proteins with unique function or for those already in use. In such cases, synthetic consensus design offers a method to introduce desired ionizable residue mutations while avoiding detrimental effects on target affinity or enzymatic activity. The synthetic design approach described herein relied on previously evolved paratopes that may not be available for novel targets. To overcome this hurdle, proteolytic challenge of combinatorial libraries using a broad protease such as proteinase K can help to determine mutations that retain structural fold and stability.

Due to their capacity for broad functional diversity, proteins are a protean class of biomolecules that have seen significant implementation in catalytic, structural, diagnostic, and therapeutic applications. The continued discovery and evolution of proteins with novel or improved functionality would be greatly benefited by the enhanced physicochemical robustness and tunable *in vivo* properties achievable through ionic charge engineering. The methods and accomplishments described in this thesis make direct contributions towards the advancement of protein discovery and developability through charge engineering.

# Bibliography


1. Akoh, C. C., Chang, S. W., Lee, G. C. & Shaw, J. F. Biocatalysis for the production of industrial products and functional foods from rice and other agricultural produce. *Journal of Agricultural and Food Chemistry* **56,** 10445–10451 (2008).
2. Gupta, R., Beg, Q. & Lorenz, P. Bacterial alkaline proteases: Molecular approaches and industrial applications. *Applied Microbiology and Biotechnology* **59,** 15–32 (2002).
3. Kirk, O., Borchert, T. V. & Fuglsang, C. C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13,** 345–351 (2002).
4. Lei, X. G. & Stahl, C. H. Nutritional benefits of phytase and dietary determinants of its efficacy. *J. Appl. Anim. Res.* **17,** 97–112 (2000).
5. de Souza, P. M. & e Magalhães, P. de O. Application of microbial α-amylase in industry - a review. *Brazilian J. Microbiol.* **41,** 850–861 (2010).
6. Pantoliano, M. W. *et al.* Large Increases in General Stability for Subtilisin BPN' through Incremental Changes in the Free Energy of Unfolding. *Biochemistry* **28,** 7205–7213 (1989).
7. Declerck, N., Machius, M., Wiegand, G., Huber, R. & Gaillardin, C. Probing structural determinants specifying high thermostability in Bacillus licheniformis α-amylase. *J. Mol. Biol.* **301,** 1041–1057 (2000).
8. Lehmann, M. *et al.* From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. *Protein Eng. Des. Sel.* **13,** 49–57 (2000).
9. Shaw, A., Bott, R. & Day, A. G. Protein engineering of α-amylase for low pH performance. *Curr. Opin. Biotechnol.* **10,** 349–352 (1999).
10. Bassegoda, A., Cesarini, S. & Diaz, P. Lipase Improvement: Goals and Strategies. *Comput. Struct. Biotechnol. J.* **2,** e201209005 (2012).
11. Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **67,** 509–544 (1998).
12. Chudakov, D. M., Lukyanov, S. & Lukyanov, K. A. Fluorescent proteins as a toolkit for in vivo imaging. *Trends Biotechnol.* **23,** 605–613 (2005).
13. Kranaster, R. & Marx, A. Engineered DNA polymerases in biotechnology. *ChemBioChem* **11,** 2077–2084 (2010).
14. Lohman, G. J. S., Tabor, S. & Nichols, N. M. DNA Ligases. *Curr. Protoc. Mol. Biol.* **2011,** 10–16 (2011).
15. Beyzavi, K. *et al.* Comparison of horseradish peroxidase and alkaline phosphatase-labelled antibodies in enzyme immunoassays. *Ann. Clin. Biochem.* **24,** 145–152 (1987).
16. Saxena, R. & Nanjan, M. J. Elastin-like polypeptides and their applications in anticancer drug delivery systems: A review. *Drug Deliv.* **22,** 156–167 (2015).
17. Shen, W., Zhang, K., Kornfield, J. A. & Tirrell, D. A. Tuning the erosion rate of artificial protein hydrogels through control of network topology. *Nat. Mater.* **5,** 153–158 (2006).
18. Schacht, K. & Scheibel, T. Controlled hydrogel formation of a recombinant spider silk protein. *Biomacromolecules* **12,** 2488–2495 (2011).
19. Kapoor, S. & Kundu, S. C. Silk protein-based hydrogels: Promising advanced materials for biomedical applications. *Acta Biomaterialia* **31,** 17–32 (2016).
20. Malanovic, N. & Lohner, K. *Antimicrobial peptides targeting Gram-positive bacteria*.

*Pharmaceuticals* **9,** (2016).
21. Volzing, K., Borrero, J., Sadowsky, M. J. & Kaznessis, Y. N. Antimicrobial peptides targeting gram-negative pathogens, produced and delivered by lactic acid bacteria. *ACS Synth. Biol.* **2,** 643–650 (2013).
22. Da Silva, A. & Teschke, O. Effects of the antimicrobial peptide PGLa on live Escherichia coli. *Biochim. Biophys. Acta - Mol. Cell Res.* **1643,** 95–103 (2003).
23. Vigant, F., Santos, N. C. & Lee, B. Broad-spectrum antivirals against viral fusion. *Nat. Rev. Microbiol.* **13,** 426–437 (2015).
24. Wu, W. *et al.* New influenza a virus entry inhibitors derived from the viral fusion peptides. *PLoS One* **10,** e0138426 (2015).
25. Freire, J. M. *et al.* siRNA-cell-penetrating peptides complexes as a combinatorial therapy against chronic myeloid leukemia using BV173 cell line as model. *J. Control. Release* **245,** 127–136 (2017).
26. Rodrigues, M., De La Torre, B. G., Rádis-Baptista, G., Santos, N. C. & Andreu, D. Efficient cellular delivery of β-galactosidase mediated by NrTPs, a new family of cell-penetrating peptides. *Bioconjug. Chem.* **22,** 2339–2344 (2011).
27. Jonker, D. J. *et al.* Cetuximab for the Treatment of Colorectal Cancer. *N. Engl. J. Med.* **357,** 2040–2048 (2007).
28. Baselga, J. Clinical trials of Herceptin® (trastuzumab). *Eur. J. Cancer* **37,** 18–24 (2001).
29. Ross, S. L. *et al.* Bispecific T cell engager (BiTE®) antibody constructs can mediate bystander tumor cell killing. *PLoS One* **12,** 1–24 (2017).
30. Lambert, J. M. & Morris, C. Q. Antibody–Drug Conjugates (ADCs) for Personalized Treatment of Solid Tumors: A Review. *Adv. Ther.* **34,** 1015–1035 (2017).
31. Ha, S. Y. *et al.* MET overexpression assessed by new interpretation method predicts gene amplification and poor survival in advanced gastric carcinomas. *Mod. Pathol.* **26,** 1632–1641 (2013).
32. Bang, Y. J. *et al.* Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): A phase 3, open-label, randomised controlled trial. *Lancet* **376,** 687–697 (2010).
33. Painter, J. T., Clayton, N. P. & Herbert, R. A. Useful immunohistochemical markers of tumor differentiation. *Toxicologic Pathology* **38,** 131–141 (2010).
34. Scartozzi, M. *et al.* Epidermal growth factor receptor (EGFR) status in primary colorectal tumors does not correlate with EGFR expression in related metastatic sites: Implications for treatment with EGFR-targeted monoclonal antibodies. *J. Clin. Oncol.* **22,** 4720–4726 (2004).
35. Van den Burg, B. & Eijsink, V. G. Selection of mutations for increased protein stability. *Current Opinion in Biotechnology* **13,** 333–337 (2002).
36. Schmidt, M. M. & Wittrup, K. D. A modeling analysis of the effects of molecular size and binding affinity on tumor targeting. *Mol. Cancer Ther.* **8,** 2861–2871 (2009).
37. Mosavi, L. K. & Peng, Z. -y. Structure-based substitutions for increased solubility of a designed protein. *Protein Eng. Des. Sel.* **16,** 739–745 (2003).
38. Hackel, B. J., Sathirachinda, A. & Gambhir, S. S. Designed hydrophilic and charge mutations of the fibronectin domain: Towards tailored protein biodistribution. *Protein Eng. Des. Sel.* **25,** 639–647 (2012).
39. Pimm, M. V., Gribben, S. J., Bogdan, K. & Hudecz, F. The effect of charge on the

biodistribution in mice of branched polypeptides with a poly(L-lysine) backbone labelled with 1251, 111In or 51Cr. *J. Control. Release* **37,** 161–172 (1995).
40. Bumbaca, D. *et al.* Highly specific off-target binding identified and eliminated during the humanization of an antibody against FGF receptor 4. *MAbs* **3,** 376–386 (2011).
41. Yang, K. *et al.* Tailoring structure-function and pharmacokinetic properties of single-chain Fv proteins by site-specific PEGylation. *Protein Eng. Des. Sel.* **16,** 761–770 (2003).
42. Chapman, A. P. *et al.* Therapeutic antibody fragments with prolonged in vivo half-lives. *Nat. Biotechnol.* **17,** 780–783 (1999).
43. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci.* **103,** 5869–5874 (2006).
44. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4,** 35–37 (2008).
45. Bloom, J. D. *et al.* Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci.* **102,** 606–611 (2005).
46. Markert, Y., Köditz, J., Mansfeld, J., Arnold, U. & Ulbrich-Hofmann, R. Increased proteolytic resistance of ribonuclease A by protein engineering. *Protein Eng. Des. Sel.* **14,** 791–796 (2001).
47. Frenken, L. G., Egmond, M. R., Batenburg, A. M. & Verrips, C. T. Pseudomonas glumae lipase: Increased proteolytic stability by protein engineering. *Protein Eng. Des. Sel.* **6,** 637–642 (1993).
48. Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **99,** 303–310 (2005).
49. Onoue, S. *et al.* Mishandling of the Therapeutic Peptide Glucagon Generates Cytotoxic Amyloidogenic Fibrils Satomi. *Pharm. Res.* **21,** 1274–1283 (2004).
50. Rosenberg, A. S. Effects of protein aggregates: An immunologic perspective. *AAPS J.* **8,** E501–E507 (2006).
51. Chan, W. *et al.* Mutational effects on inclusion body formation in the periplasmic expression of the immunoglobulin VLdomain REI. *Fold. Des.* **1,** 77–89 (1996).
52. Yan, G. *et al.* A single residual replacement improves the folding and stability of recombinant cassava hydroxynitrile lyase in E. coli. *Biotechnol. Lett.* **25,** 1041–1047 (2003).
53. Hashimoto, K. & Panchenko, A. R. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci.* **107,** 20352–20357 (2010).
54. Li, Z., Zhang, J., Jin, Z., Zhang, W. & Zhang, Y. Synthesis and biodistribution of novel 99m Tc labeled 4-nitroimidazole dithiocarbamate complexes as potential agents to target tumor hypoxia. *Medchemcomm* **6,** 1143–1148 (2015).
55. Ono, M. *et al.* Control of radioactivity pharmacokinetics of 99mTc-HYNIC-labeled polypeptides derivatized with ternary ligand complexes. *Bioconjug. Chem.* **13,** 491–501 (2002).
56. Isom, D. G., Castaneda, C. A., Cannon, B. R., Velu, P. D. & Garcia-Moreno E., B. Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci.* **107,** 16096–16100 (2010).
57. Schwehm, J. M., Fitch, C. A., Dang, B. N., Bertrand García-Moreno, E. & Stites, W. E. Changes in stability upon charge reversal and neutralization substitution in staphylococcal nuclease are dominated by favorable electrostatic effects. *Biochemistry* **42,** 1118–1128 (2003).

58. Dao-pin, S., Anderson, D. E., Baase, W. A., Dahlquist, F. W. & Matthews, B. W. Structural and Thermodynamic Consequeces of Burying a Charged Residue within the Hydrophobic Core of T4 Lysozyme. *Biochemistry* **30,** 11521–11528 (1991).
59. Perutz, M. F., Kendrew, J. C. & Watson, H. C. Structure and Function of Haemoglobin n. Some Relations between Polypeptide Chain Configuration and. *J. Mol. Biol.* **13,** 669–678 (1965).
60. Zhou, H.-X. & Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* acs.chemrev.7b00305 (2018). doi:10.1021/acs.chemrev.7b00305
61. Cao, Z. & Bowie, J. U. An energetic scale for equilibrium H/D fractionation factors illuminates hydrogen bond free energies in proteins. *Protein Sci.* **23,** 566–575 (2014).
62. Nick Pace, C., Alston, R. W. & Shaw, K. L. Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci.* **9,** 1395–1398 (2000).
63. Grimsley, G. R. *et al.* Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8,** 1843–9 (1999).
64. Pace, C. N., Grimsley, G. R. & Scholtz, J. M. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J. Biol. Chem.* **284,** 13285–13289 (2009).
65. Kumar, S. & Nussinov, R. Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **293,** 1241–1255 (1999).
66. Novák, P. & Havlíček, V. *Proteomic Profiling and Analytical Chemistry*. *Proteomic Profiling and Analytical Chemistry* (2016). doi:10.1016/B978-0-444-63688-1.00004-5
67. Chan, P., Curtis, R. A. & Warwicker, J. Soluble expression of proteins correlates with a lack of positively-charged surface. *Sci. Rep.* **3,** 3333 (2013).
68. Kim, J., Mosior, M., Chung, L. A., Wu, H. & McLaughlin, S. Binding of peptides with basic residues to membranes containing acidic phospholipids. *Biophys. J.* **60,** 135–148 (1991).
69. Gross, W. *Biological Membranes*. *Angewandte Chemie International Edition in English* **10,** (1971).
70. Murray, D. *et al.* Electrostatic properties of membranes containing acidic lipids and adsorbed basic peptides: Theory and experiment. *Biophys. J.* **77,** 3176–3188 (1999).
71. Boswell, C. A. *et al.* Effects of charge on antibody tissue distribution and pharmacokinetics. *Bioconjugate Chemistry* **21,** 2153–2163 (2010).
72. Kangas, E. & Tidor, B. Charge optimization leads to favorable electrostatic binding free energy. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **59,** 5958–5961 (1999).
73. Traxlmayr, M. W. *et al.* Strong enrichment of aromatic residues in binding sites from a charge-neutralized hyperthermostable Sso7D scaffold library. *J. Biol. Chem.* **291,** 22496–22508 (2016).
74. Kastritis, P. L., Rodrigues, J. P. G. L. M., Folkers, G. E., Boelens, R. & Bonvin, A. M. J. J. *Proteins feel more than they see: Fine-tuning of binding affinity by properties of the non-interacting surface*. *Journal of Molecular Biology* **426,** (Elsevier B.V., 2014).
75. Dellian, M., Yuan, F., Trubetskoy, V. S., Torchilin, V. P. & Jain, R. K. Vascular permeability in a human tumour xenograft: molecular charge dependence M. *Br. J. Cancer* **83,** 177–183 (2000).
76. Wiig, H., Gyenge, C. C. & Tenstad, O. The interstitial distribution of macromolecules in rat tumours is influenced by the negatively charge matrix components. *J. Physiol.* **567,** 557–

567 (2005).
77. Krasnici, S. *et al.* Effect of the surface charge of liposomes on their uptake by angiogenic tumor vessels. *Int. J. Cancer* **105,** 561–567 (2003).
78. Stylianopoulos, T., Soteriou, K., Fukumura, D. & Jain, R. K. Cationic nanoparticles have superior transvascular flux into solid tumors: Insights from a mathematical model. *Ann. Biomed. Eng.* **41,** 68–77 (2013).
79. Kimura, R. H. *et al.* Pharmacokinetically stabilized cystine knot peptides that bind alpha-v-beta-6 integrin with single-digit nanomolar affinities for detection of pancreatic cancer. *Clin. Cancer Res.* **18,** 839–849 (2012).
80. Tran, T. A. *et al.* Effects of lysine-containing mercaptoacetyl-based chelators on the biodistribution of 99mTc-labeled anti-HER2 affibody molecules. *Bioconjug. Chem.* **19,** 2568–2576 (2008).
81. Ekblad, T. *et al.* Development and preclinical characterisation of 99mTc-labelled Affibody molecules with reduced renal uptake. *Eur. J. Nucl. Med. Mol. Imaging* **35,** 2245–2255 (2008).
82. García Garayoa, E. *et al.* Influence of the molecular charge on the biodistribution of bombesin analogues labeled with the [99mTc(CO)3]-core. *Bioconjug. Chem.* **19,** 2409–2416 (2008).
83. Ma, S. F., Nishikawa, M., Katsumi, H., Yamashita, F. & Hashida, M. Cationic charge-dependent hepatic delivery of amidated serum albumin. *J. Control. Release* **102,** 583–594 (2005).
84. Kobayashi, H. *et al.* The pharmacokinetic characteristics of glycolated humanized anti-Tac Fabs are determined by their isoelectric points. *Cancer Res.* **59,** 422–430 (1999).
85. Taverna, D. M. & Goldstein, R. A. Why are proteins marginally stable? *Proteins Struct. Funct. Genet.* **46,** 105–109 (2002).
86. Pace, C. N. & Hermans, J. The Stability of Globular Protein. *CRC Crit. Rev. Biochem.* **3,** 1–43 (1975).
87. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology* **19,** 596–604 (2009).
88. Guo, H. H., Choe, J. & Loeb, L. A. Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci.* **101,** 9205–9210 (2004).
89. Poluri, K. M. & Gulati, K. *Protein engineering techniques: Gateways to synthetic protein universe*. *SpringerBriefs in Applied Sciences and Technology* (2017). doi:10.1007/978-981-10-2732-1
90. Trevino, S. R., Scholtz, J. M. & Pace, C. N. Measuring and increasing protein solubility. *Journal of Pharmaceutical Sciences* **97,** 4155–4166 (2008).
91. Trevino, S. R., Scholtz, J. M. & Pace, C. N. Amino Acid Contribution to Protein Solubility: Asp, Glu, and Ser Contribute more Favorably than the other Hydrophilic Amino Acids in RNase Sa. *J. Mol. Biol.* **366,** 449–460 (2007).
92. Raghunathan, G. *et al.* Modulation of protein stability and aggregation properties by surface charge engineering. *Mol. Biosyst.* **9,** 2379 (2013).
93. Pédelacq, J.-D. *et al.* Engineering soluble proteins for structural genomics. *Nat. Biotechnol.* **20,** 927–932 (2002).
94. Hackel, B. B. J. B. B. J., Kimura, R. R. H. & Gambhir, S. S. S. Use of 64 Cu-labeled Fibronectin Domain with EGFR-Overexpressing Tumor Xenograft: Molecular Imaging. *Radiology* **263,** 179–188 (2012).

95. Onda, M., Kreitman, R. J., Vasmatzis, G., Lee, B. & Pastan, I. Reduction of the nonspecific animal toxicity of anti-Tac(Fv)-PE38 by mutations in the framework regions of the Fv which lower the isoelectric point. *J. Immunol.* **163,** 6072–7 (1999).
96. Hofström, C. *et al.* Use of a HEHEHE purification tag instead of a hexahistidine tag improves biodistribution of affibody molecules site-specifically labeled with 99mTc, 111In, and 125I. *J. Med. Chem.* **54,** 3817–3826 (2011).
97. Lindbo, S. *et al.* Influence of Histidine-Containing Tags on the Biodistribution of ADAPT Scaffold Proteins. *Bioconjug. Chem.* **27,** 716–726 (2016).
98. Chen, J., Sawyer, N. & Regan, L. Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci.* **22,** 510–515 (2013).
99. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA. Cancer J. Clin.* **68,** 7–30 (2018).
100. American Cancer Society. Cancer Facts & Figures 2018.
101. Howlader, N. *et al.* SEER Cancer Statistics Review, 1975-2014. *Natl. Cancer Institute. Bethesda, MD* (2017).
102. Malhotra, V. & Perry, M. C. Classical chemotherapy: mechanisms, toxicities and the therapeutic window. *Cancer Biol. Ther.* **2,** 4–6 (2003).
103. Etzioni, R. *et al.* Early detection: The case for early detection. *Nat. Rev. Cancer* **3,** 243–252 (2003).
104. Hayes, D. F. *et al.* HER2 and Response to Paclitaxel in Node-Positive Breast Cancer. *N. Engl. J. Med.* **357,** 1496–1506 (2007).
105. Van Cutsem, E. *et al.* Bevacizumab in Combination With Chemotherapy As First-Line Therapy in Advanced Gastric Cancer: A Biomarker Evaluation From the AVAGAST Randomized Phase III Trial. *J. Clin. Oncol.* **30,** 2119–2127 (2012).
106. Korpanty, G. J., Graham, D. M., Vincent, M. D. & Leighl, N. B. Biomarkers That Currently Affect Clinical Practice in Lung Cancer: EGFR, ALK, MET, ROS-1, and KRAS. *Front. Oncol.* **4,** 1–8 (2014).
107. Herbst, R. S. Review of epidermal growth factor receptor biology. *Int. J. Radiat. Oncol. Biol. Phys.* **59,** 21–26 (2004).
108. Wee, P. & Wang, Z. Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers (Basel).* **9,** (2017).
109. Nicholson, R. ., Gee, J. M. . & Harper, M. . EGFR and cancer prognosis. *Eur. J. Cancer* **37,** 9–15 (2001).
110. Laurent-Puig, P. *et al.* Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. *J. Clin. Oncol.* **27,** 5924–5930 (2009).
111. Scartozzi, M. *et al.* Epidermal Growth Factor Receptor (EGFR) gene copy number (GCN) correlates with clinical activity of irinotecan-cetuximab in K-RAS wild-type colorectal cancer: A fluorescence in situ (FISH) and chromogenic in situ hybridization (CISH) analysis. *BMC Cancer* **9,** 303 (2009).
112. Pirker, R. *et al.* EGFR expression as a predictor of survival for first-line chemotherapy plus cetuximab in patients with advanced non-small-cell lung cancer: Analysis of data from the phase 3 FLEX study. *Lancet Oncol.* **13,** 33–42 (2012).
113. Douillard, J.-Y. *et al.* Relationship Between EGFR Expression, EGFR Mutation Status, and the Efficacy of Chemotherapy Plus Cetuximab in FLEX Study Patients with Advanced Non–

Small-Cell Lung Cancer. *J. Thorac. Oncol.* **9,** 717–724 (2014).
114. Huang, J. *et al.* Icotinib in patients with pretreated advanced esophageal squamous cell Carcinoma with EGFR overexpression or EGFR gene amplification: A single-arm, multicenter phase 2 study. *J. Thorac. Oncol.* **11,** 910–917 (2016).
115. Lee, H. J. *et al.* Prognostic and predictive values of EGFR overexpression and EGFR copy number alteration in HER2-positive breast cancer. *Br. J. Cancer* **112,** 103–111 (2015).
116. Huang, L. *et al.* SPECT Imaging with 99mTc-Labeled EGFR-Specific Nanobody for In Vivo Monitoring of EGFR Expression. *Mol. Imaging Biol.* **10,** 167–175 (2008).
117. Miao, Z. *et al.* PET of EGFR Expression with an 18F-Labeled Affibody Molecule. *J. Nucl. Med.* **53,** 1110–1118 (2012).
118. Zidan, J. *et al.* Comparison of HER-2 overexpression in primary breast cancer and metastatic sites and its effect on biological targeting therapy of metastatic disease. *Br. J. Cancer* **93,** 552–556 (2005).
119. Jensen, J. D. *et al.* PIK3CA mutations may be discordant between primary and corresponding metastatic disease in breast cancer. *Clin. Cancer Res.* **17,** 667–677 (2011).
120. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366,** 883–892 (2012).
121. Kobayashi, S. *et al.* EGFR Mutation and Resistance of Non–Small-Cell Lung Cancer to Gefitinib. *N. Engl. J. Med.* **352,** 786–792 (2005).
122. Pacak, J., Tocik, Z. & Cerny, M. Synthesis of 2-Deoxy-2-Fluoro-D-Glucose. *Chem. Commun.* **1454,** 77–77 (1969).
123. Roosenburg, S. *et al.* PET and SPECT imaging of a radiolabeled minigastrin analogue conjugated with DOTA, NOTA, and NODAGA and labeled with 64Cu, 68Ga, and 111In. *Mol. Pharm.* **11,** 3930–3937 (2014).
124. Alva-Sanchez, H., Quintana-Bautista, C., Martïnez-Davalos, A., Avila-Rodrïguez, M. A. & Rodrïguez-Villafuerte, M. Positron range in tissue-equivalent materials: Experimental microPET studies. *Phys. Med. Biol.* **61,** 6307–6321 (2016).
125. Jødal, L., Le Loirec, C. & Champion, C. Positron range in PET imaging: Non-conventional isotopes. *Phys. Med. Biol.* **59,** 7419–7434 (2014).
126. Conti, M. & Eriksson, L. Physics of pure and non-pure positron emitters for PET: A review and a discussion. *EJNMMI Phys.* **3,** (2016).
127. Wittrup, K. D., Thurber, G. M., Schmidt, M. M. & Rhoden, J. J. Practical theoretic guidance for the design of tumor-targeting agents. *Methods Enzymol.* **503,** 255–268 (2012).
128. Orlova, A. *et al.* On the selection of a tracer for PET imaging of HER2-expressing tumors: direct comparison of a 124I-labeled affibody molecule and trastuzumab in a murine xenograft model. *J. Nucl. Med.* **50,** 417–425 (2009).
129. Zahnd, C. *et al.* Efficient tumor targeting with high-affinity designed ankyrin repeat proteins: Effects of affinity and molecular size. *Cancer Res.* **70,** 1595–1605 (2010).
130. Natarajan, A., Hackel, B. J. & Gambhir, S. S. A Novel Engineered Anti-CD20 Tracer Enables Early Time PET Imaging in a Humanized Transgenic Mouse Model of B-cell Non-Hodgkins Lymphoma. *Clin. Cancer Res.* **19,** 6820–6829 (2013).
131. Stern, L., Case, B. & Hackel, B. Alternative Non-Antibody Scaffolds for Molecular Imaging of Cancer. *Curr. Opin. Chem. Eng.* **2,** 425–432 (2013).
132. Vazquez-Lombardi, R. *et al.* Challenges and opportunities for non-antibody scaffold drugs. *Drug Discov. Today* **20,** 1271–1283 (2015).
133. Wahlberg, E. *et al.* An affibody in complex with a target protein: structure and coupled

folding. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 3185–3190 (2003).
134. Camara, B. *et al.* T7 phage protein Gp2 inhibits the Escherichia coli RNA polymerase by antagonizing stable DNA strand separation near the transcription start site. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2247–52 (2010).
135. Arora, P. Fast and faster: A designed variant of the B-domain of protein A folds in 3 μsec. *Protein Sci.* **13,** 847–853 (2004).
136. Nilsson, B. *et al.* A synthetic IgG-binding domain based on staphylococcal protein a. *Protein Eng. Des. Sel.* **1,** 107–113 (1987).
137. Löfblom, J. *et al.* Affibody molecules: Engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS Lett.* **584,** 2670–2680 (2010).
138. Ståhl, S. *et al.* Affibody Molecules in Biotechnological and Medical Applications. *Trends in Biotechnology* **35,** 691–712 (2017).
139. Sandstrom, M. *et al.* Biodistribution and Radiation Dosimetry of the Anti-HER2 Affibody Molecule 68Ga-ABY-025 in Breast Cancer Patients. *J. Nucl. Med.* **57,** 867–871 (2016).
140. Sörensen, J. *et al.* Measuring HER2-receptor expression in metastatic breast cancer using [68Ga]ABY-025 Affibody PET/CT. *Theranostics* **6,** 262–271 (2016).
141. Sorensen, J. *et al.* First-in-Human Molecular Imaging of HER2 Expression in Breast Cancer Metastases Using the 111In-ABY-025 Affibody Molecule. *J. Nucl. Med.* **55,** 730–735 (2014).
142. In Vivo IGF-1R Molecular Imaging Using [68Ga]- Labelling Anti-IGF-1R Affibody Molecule. Available at: https://clinicaltrials.gov/ct2/show/NCT02916394?term=affibody&rank=2. (Accessed: 5th April 2018)
143. In Vivo EGFR-ECD Molecular Imaging Using [68Ga]-Labelling Anti-EGFR Affibody Molecule. Available at: https://clinicaltrials.gov/ct2/show/NCT02916329?term=affibody&rank=1. (Accessed: 5th April 2018)
144. Feldwisch, J. *et al.* Design of an optimized scaffold for affibody molecules. *J. Mol. Biol.* **398,** 232–47 (2010).
145. Engfeldt, T. *et al.* 99mTc-chelator engineering to improve tumour targeting properties of a HER2-specific Affibody molecule. *Eur. J. Nucl. Med. Mol. Imaging* **34,** 1843–1853 (2007).
146. Wallberg, H. *et al.* Molecular Design and Optimization of 99mTc-Labeled Recombinant Affibody Molecules Improves Their Biodistribution and Imaging Properties. *J. Nucl. Med.* **52,** 461–469 (2011).
147. Webster, J. M., Zhang, R., Gambhir, S. S., Cheng, Z. & Syud, F. A. Engineered two-helix small proteins for molecular recognition. *ChemBioChem* **10,** 1293–1296 (2009).
148. Ren, G. *et al.* A 2-helix small protein labeled with 68Ga for PET imaging of HER2 expression. *J. Nucl. Med.* **50,** 1492–1499 (2009).
149. Kruziki, M. A. A., Bhatnagar, S., Woldring, D. R. R., Duong, V. T. T. & Hackel, B. J. J. A 45-Amino-Acid Scaffold Mined from the PDB for High-Affinity Ligand Engineering. *Chem. Biol.* **22,** 946–956 (2015).
150. Chan, J. Y., Hackel, B. J. & Yee, D. Targeting Insulin Receptor in Breast Cancer Using Small Engineered Protein Scaffolds. *Mol. Cancer Ther.* **16,** 1324–1334 (2017).
151. Slovic, A. M., Kono, H., Lear, J. D., Saven, J. G. & DeGrado, W. F. Computational design of water-soluble analogues of the potassium channel KcsA. *Proc. Natl. Acad. Sci.* **101,** 1828–1833 (2004).
152. Jenkins, T. M. *et al.* Catalytic domain of human immunodeficiency virus type 1 integrase:

identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci U S A* **92,** 6057–6061 (1995).
153. Zhang, F. *et al.* Crystal structure of the obese protein leptin-E100. *Nature* **387,** 206–209 (1997).
154. Pimm, M. V. *et al.* Gamma scintigraphy of111In-labelled branched chain polypeptides (BCP) with a poly(L-lysine) backbone in mice with mammary carcinoma: Effect of charge on biodistribution and tumour imaging potential. *Ann. Nucl. Med.* **9,** 247–251 (1995).
155. Tran, T. *et al.* 99mTc-maEEE-ZHER2:342, an affibody molecule-based tracer for the detection of HER2 expression in malignant tumors. *Bioconjug. Chem.* **18,** 1956–1964 (2007).
156. Dale, G. E., Broger, C., Langen, H., Arcy, A. D. & Stüber, D. Improving protein solubility through rationally designed amino acid replacements: Solubilization of the trimethoprim-resistant type s1 dihydrofolate reductase. *Protein Eng. Des. Sel.* **7,** 933–939 (1994).
157. Das, D. & Georgiadis, M. M. A directed approach to improving the solubility of Moloney murine leukemia virus reverse transcriptase. *Protein Sci.* **10,** 1936–41 (2001).
158. Igawa, T. *et al.* Reduced elimination of IgG antibodies by engineering the variable region. *Protein Eng. Des. Sel.* **23,** 385–392 (2010).
159. Nygren, P. Å. Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS J.* **275,** 2668–2676 (2008).
160. Friedman, M. *et al.* Directed Evolution to Low Nanomolar Affinity of a Tumor-Targeting Epidermal Growth Factor Receptor-Binding Affibody Molecule. *J. Mol. Biol.* **376,** 1388–1402 (2008).
161. Hynes, N. E. & MacDonald, G. ErbB receptors and signaling pathways in cancer. *Curr. Opin. Cell Biol.* **21,** 177–184 (2009).
162. Steipe, B. *et al.* Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* **240,** 188–192 (1994).
163. Steipe, B. Consensus-based engineering of protein stability: From intrabodies to thermostable enzymes. *Methods in Enzymology* **388,** 176–186 (2004).
164. Cochran, J. R., Kim, Y. S., Lippow, S. M., Rao, B. & Wittrup, K. D. Improved mutants from directed evolution are biased to orthologous substitutions. *Protein Eng. Des. Sel.* **19,** 245–253 (2006).
165. Durani, V. & Magliery, T. J. Protein engineering and stabilization from sequence statistics: Variation and covariation analysis. *Methods Enzymol.* **523,** 237–256 (2013).
166. Sullivan, B. J., Durani, V. & Magliery, T. J. Triosephosphate isomerase by consensus design: Dramatic differences in physical properties and activity of related variants. *J. Mol. Biol.* **413,** 195–208 (2011).
167. Boder, E. T. & Wittrup, K. D. Yeast Surface Display for Screening Combinatorial Polypeptide Libraries. *Nat. Biotechnol.* **15,** 553–557 (1997).
168. Jäckel, C., Bloom, J. D., Kast, P., Arnold, F. H. & Hilvert, D. Consensus protein design without phylogenetic bias. *J. Mol. Biol.* **399,** 541–546 (2010).
169. Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Res.* **42,** D211-22 (2014).
170. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89,** 10915–10919 (1992).
171. Pogulis, R. J., Vallejo, A. N. & Pease, L. R. In Vitro Recombination and Mutagenesis by Overlap Extension PCR. *Vitr. Mutagen. Protoc.* **57,** 167–176 (1996).

172. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1,** 755–768 (2006).
173. Spangler, J. B. *et al.* Combination antibody treatment down-regulates epidermal growth factor receptor by inhibiting endosomal recycling. *Proc. Natl. Acad. Sci.* **107,** 13252–13257 (2010).
174. Orlova, A. *et al.* Tumor imaging using a picomolar affinity HER2 binding Affibody molecule. *Cancer Res.* **66,** 4339–4348 (2006).
175. Shannon, C. E. & E., C. A mathematical theory of communication. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **5,** 3 (2001).
176. Patrick, W. M., Firth, A. E. & Blackburn, J. M. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng. Des. Sel.* **16,** 451–457 (2003).
177. Pavoor, T. V., Wheasler, J. A., Kamat, V. & Shusta, E. V. An enhanced approach for engineering thermally stable proteins using yeast display. *Protein Eng. Des. Sel.* **25,** 625–630 (2012).
178. Malm, M. *et al.* Inhibiting HER3-Mediated Tumor Cell Growth with Affibody Molecules Engineered to Low Picomolar Affinity by Position-Directed Error-Prone PCR-Like Diversification. *PLoS One* **8,** (2013).
179. James, M. L. & Gambhir, S. S. A molecular imaging primer: modalities, imaging agents, and applications. *Physiol. Rev.* **92,** 897–965 (2012).
180. Gebauer, M. & Skerra, A. Engineered protein scaffolds as next-generation antibody therapeutics. *Curr. Opin. Chem. Biol.* **13,** 245–255 (2009).
181. Banta, S., Dooley, K. & Shur, O. Replacing Antibodies: Engineering New Binding Proteins. *Annu. Rev. Biomed. Eng.* **15,** 93–113 (2013).
182. Straw, S., Ferrigno, P. K., Song, Q., Tomlinson, D. & Galdo, F. Del. Proof of concept study to identify candidate biomarkers of fibrosis using high throughput peptide aptamer microarray and validate by enzyme linked immunosorbant assay. *J. Biomed. Sci. Eng.* **6,** 32–42 (2013).
183. Johnson, A., Song, Q., Ko Ferrigno, P., Bueno, P. R. & Davis, J. J. Sensitive affimer and antibody based impedimetric label-free assays for C-reactive protein. *Anal. Chem.* **84,** 6553–6560 (2012).
184. Škrlec, K., Štrukelj, B. & Berlec, A. Non-immunoglobulin scaffolds: A focus on their targets. *Trends Biotechnol.* **33,** 408–418 (2015).
185. Holliger, P. & Hudson, P. J. Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* **23,** 1126–1136 (2005).
186. Rosik, D. *et al.* Incorporation of a triglutamyl spacer improves the biodistribution of synthetic affibody molecules radiofluorinated at the n-terminus via oxime formation with 18F-4-fluorobenzaldehyde. *Bioconjug. Chem.* **25,** 82–92 (2014).
187. Whitlow, M. *et al.* An improved linker for single-chain fv with reduced aggregation and enhanced proteolytic stability. *Protein Eng. Des. Sel.* **6,** 989–995 (1993).
188. Willuda, J. *et al.* High thermal stability is essential for tumor targeting of antibody fragments: Engineering of a humanized anti-epithelial glycoprotein-2 (epithelial cell adhesion molecule) single-chain Fv fragment. *Cancer Res.* **59,** 5758–5767 (1999).
189. Tran, T. *et al.* Detection of HER2 Expression in Malignant Tumors. *Tumor Target.* **2,** 1956–1964 (2007).
190. Hong, G., Bazin-Redureau, M. I. & Scherrmann, J. M. G. Pharmacokinetics and organ

distribution of cationized colchicine- specific IgG and Fab fragments in rat. *J. Pharm. Sci.* **88,** 147–153 (1999).
191. Tabata, Y., Kawai, T., Murakami, Y. & Ikada, Y. Electric charge influence of dextran derivatives on their tumor accumulation after intravenous injection. *Drug Deliv. J. Deliv. Target. Ther. Agents* **4,** 213–221 (1997).
192. Thurber, G. M., Schmidt, M. M. & Wittrup, K. D. Factors determining antibody distribution in tumors. *Trends Pharmacol. Sci.* **29,** 57–61 (2008).
193. c. Nick, P. Measuring and increasing protein stability. *Tibtech* **8,** 93–98 (1990).
194. Nord, K. *et al.* Binding proteins selected from combinatorial libraries of an α-helical bacterial receptor domain. *Nat. Biotechnol.* **15,** 772–777 (1997).
195. Grönwall, C. *et al.* Affibody-mediated transferrin depletion for proteomics applications. *Biotechnol. J.* **2,** 1389–1398 (2007).
196. Löfdahl, P.-A., Nord, O., Janzon, L. & Nygren, P.-A. Selection of TNF-alpha binding affibody molecules using a beta-lactamase protein fragment complementation assay. *N. Biotechnol.* **26,** 251–259 (2009).
197. Wikman, M. *et al.* Selection and characterization of an HIV-1 gp120-binding affibody ligand. *Biotechnol. Appl. Biochem.* **45,** 93 (2006).
198. Baum, R. P. *et al.* Molecular Imaging of HER2-Expressing Malignant Tumors in Breast Cancer Patients Using Synthetic 111In- or 68Ga-Labeled Affibody Molecules. *J. Nucl. Med.* **51,** 892–897 (2010).
199. Tolmachev, V., Velikyan, I., Sandström, M. & Orlova, A. A HER2-binding Affibody molecule labelled with 68Ga for PET imaging: direct in vivo comparison with the 111In-labelled analogue. *Eur. J. Nucl. Med. Mol. Imaging* **37,** 1356–67 (2010).
200. Zielinski, R. *et al.* Affitoxin-A novel recombinant, HER2-specific, anticancer agent for targeted therapy of HER2-positive tumors. *J. Immunother.* **32,** 817–825 (2009).
201. Orlova, A., Tran, T. A., Ekblad, T., Karlström, A. E. & Tolmachev, V. 186Re-maSGS-ZHER2:342, a potential Affibody conjugate for systemic therapy of HER2-expressing tumours. *Eur. J. Nucl. Med. Mol. Imaging* **37,** 260–269 (2010).
202. Lyakhov, I. *et al.* HER2- and EGFR-specific affiprobes: Novel recombinant optical probes for cell imaging. *ChemBioChem* **11,** 345–350 (2010).
203. Nord, K., Gunneriusson, E., Uhlén, M. & Nygren, P. A. Ligands selected from combinatorial libraries of protein A for use in affinity capture of apolipoprotein A-1(M) and Taq DNA polymerase. *J. Biotechnol.* **80,** 45–54 (2000).
204. Case, B. A. & Hackel, B. J. Synthetic and natural consensus design for engineering charge within an affibody targeting epidermal growth factor receptor. *Biotechnol. Bioeng.* **113,** 1628–1638 (2016).
205. Kim, Y.-S., Bhandari, R., Cochran, J. R., Kuriyan, J. & Wittrup, K. D. Directed evolution of the epidermal growth factor receptor extracellular domain for expression in yeast. *Proteins Struct. Funct. Bioinforma.* **62,** 1026–1035 (2005).
206. Yarden, Y. The EGFR family and its ligands in human cancer. *Eur. J. Cancer* **37,** 3–8 (2001).
207. Yewale, C., Baradia, D., Vhora, I., Patil, S. & Misra, A. Epidermal growth factor receptor targeting in cancer: A review of trends and strategies. *Biomaterials* **34,** 8690–8707 (2013).
208. Moroni, M. *et al.* Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: A cohort study. *Lancet Oncol.* **6,** 279–286 (2005).

209. Hirsch, F. R. *et al.* Combination of EGFR gene copy number and protein expression predicts outcome for advanced non-small-cell lung cancer patients treated with gefitinib. *Ann. Oncol.* **18,** 752–760 (2007).
210. Nordberg, E. *et al.* In vivo and in vitro uptake of $^{111}$In, delivered with the affibody molecule (Z<inf>EGFR:955</inf>)<inf>2</inf>, in EGFR expressing tumour cells. *Oncol. Rep.* **19,** 853–857 (2008).
211. Tolmachev, V. *et al.* Imaging of EGFR expression in murine xenografts using site-specifically labelled anti-EGFR 111In-DOTA-Z EGFR:2377 Affibody molecule: aspect of the injected tracer amount. *Eur. J. Nucl. Med. Mol. Imaging* **37,** 613–622 (2010).
212. Garousi, J. *et al.* PET imaging of epidermal growth factor receptor expression in tumours using 89Zr-labelled ZEGFR:2377 affibody molecules. *Int. J. Oncol.* **48,** 1325–1332 (2016).
213. Su, X. *et al.* Comparison of two site-specifically 18F. Labeled affibodies for PET imaging of EGFR positive tumors. *Mol. Pharm.* **11,** 3947–3956 (2014).
214. Kruziki, M. A. *et al.* 64Cu-Labeled Gp2 Domain for PET Imaging of Epidermal Growth Factor Receptor. *Mol. Pharm.* **13,** 3747–3755 (2016).
215. Gainkam, L. O. T. *et al.* Comparison of the Biodistribution and Tumor Targeting of Two 99mTc-Labeled Anti-EGFR Nanobodies in Mice, Using Pinhole SPECT/Micro-CT. *J. Nucl. Med.* **49,** 788–795 (2008).
216. Ping Li, W., Meyer, L. A., Capretto, D. A., Sherman, C. D. & Anderson, C. J. Receptor-Binding, Biodistribution, and Metabolism Studies of $^{64}$ Cu-DOTA-Cetuximab, a PET-Imaging Agent for Epidermal Growth-Factor Receptor-Positive Tumors. *Cancer Biother. Radiopharm.* **23,** 158–171 (2008).
217. Cai, W. *et al.* Quantitative PET of EGFR expression in xenograft-bearing mice using64Cu-labeled cetuximab, a chimeric anti-EGFR monoclonal antibody. *Eur. J. Nucl. Med. Mol. Imaging* **34,** 850–858 (2007).
218. Bhattacharyya, S. *et al.* Zirconium-89 labeled panitumumab: A potential immuno-PET probe for HER1-expressing carcinomas. *Nucl. Med. Biol.* **40,** 451–457 (2013).
219. Chakravarty, R. *et al.* Matching the decay half-life with the biological half-life: ImmunoPET imaging with44Sc-labeled Cetuximab Fab fragment. *Bioconjug. Chem.* **25,** 2197–2204 (2014).
220. Niu, G., Li, Z., Xie, J., Le, Q.-T. & Chen, X. PET of EGFR Antibody Distribution in Head and Neck Squamous Cell Carcinoma Models. *J. Nucl. Med.* **50,** 1116–1123 (2009).
221. Memon, A. A. *et al.* Positron emission tomography (PET) imaging with [11C]-labeled erlotinib: A micro-PET study on mice with lung tumor xenografts. *Cancer Res.* **69,** 873–878 (2009).
222. Zhang, M. R. *et al.* [11C]Gefitinib ([11C]Iressa): Radiosynthesis, In Vitro uptake, and In Vivo imaging of intact murine fibrosarcoma. *Mol. Imaging Biol.* **12,** 181–191 (2010).
223. Slobbe, P. *et al.* Development of [18F]afatinib as new TKI-PET tracer for EGFR positive tumors. *Nucl. Med. Biol.* **41,** 749–757 (2014).
224. Wang, H. *et al.* Assessment of 11C-labeled-4-N-(3-bromoanilino)-6,7-dimethoxyquinazoline as a positron emission tomography agent to monitor epidermal growth factor receptor expression. *Cancer Sci.* **98,** 1413–6 (2007).
225. Velikyan, I. *et al.* Preparation and Evaluation of Ga-68-DOTA-hEGF for Visualization of EGFR Expression in Malignant Tumors. *J. Nucl. Med.* **46,** 1881–1888 (2005).
226. Kareem, H. *et al.* Blocking EGFR in the liver improves the tumor-to-liver uptake ratio of radiolabeled EGF. *Tumor Biol.* **31,** 79–87 (2010).

227. Stern, L. A. *et al.* Geometry and expression enhance enrichment of functional yeast-displayed ligands via cell panning. *Biotechnol. Bioeng.* **113,** 2328–2341 (2016).
228. Shukla, A. & Kumar, U. Positron emission tomography: An overview. *J. Med. Phys.* **31,** 13 (2006).
229. Ait-Mohand, S. *et al.* Evaluation of 64Cu-labeled bifunctional Chelate-Bombesin conjugates. *Bioconjug. Chem.* **22,** 1729–1735 (2011).
230. Miao, Z., Ren, G., Hongguang, L., Jiang, L. & Cheng, Z. Small-animal PET imaging of human epidermal growth factor receptor positive tumor with a 64Cu labeled affibody protein. *Bioconjug. Chem.* **21,** 947–954 (2010).
231. Zhao, P. *et al.* Molecular imaging of hepatocellular carcinoma xenografts with epidermal growth factor receptor targeted affibody probes. *Biomed Res. Int.* **2013,** 759057 (2013).
232. Tolmachev, V. *et al.* Radionuclide therapy of HER2-positive microxenografts using a177Lu-labeled HER2-specific affibody molecule. *Cancer Res.* **67,** 2773–2782 (2007).
233. Hanenberg, M. *et al.* Amyloid-β peptide-specific darpins as a novel class of potential therapeutics for alzheimer disease. *J. Biol. Chem.* **289,** 27080–27089 (2014).
234. Mamluk, R. *et al.* Anti-tumor effect of CT-322 as an adnectin inhibitor of vascular endothelial growth factor receptor-2. *MAbs* **2,** 199–208 (2010).
235. Dreher, M. R. *et al.* Tumor vascular permeability, accumulation, and penetration of macromolecular drug carriers. *J. Natl. Cancer Inst.* **98,** 335–344 (2006).
236. Pikarsky, E. *et al.* NF-κB functions as a tumour promoter in inflammation-associated cancer. *Nature* **431,** 461–466 (2004).
237. Thurber, G. M., Schmidt, M. M. & Wittrup, K. D. Antibody tumor penetration: Transport opposed by systemic and antigen-mediated clearance. *Adv. Drug Deliv. Rev.* **60,** 1421–1434 (2008).
238. Lobo, E. D., Hansen, R. J. & Balthasar, J. P. Antibody pharmacokinetics and pharmacodynamics. *J. Pharm. Sci.* **93,** 2645–2668 (2004).
239. Willuda, J. *et al.* High thermal stability is essential for tumor targeting of antibody fragments. *Cancer Res.* **59,** 5758–5767 (1999).
240. Chan, J. Y., Hackel, B. J. & Yee, D. Insulin receptor targeting in breast cancer through yeast surface display. *Cancer Res.* **75,** P4-6–2 (2015).
241. Pezzullo, M. *et al.* Comprehensive analysis of surface charged residues involved in thermal stability in Alicyclobacillus acidocaldarius esterase 2. *Protein Eng. Des. Sel.* **26,** 47–58 (2013).
242. Matsuura, Y. *et al.* Role of charged residues in stabilization of Pyrococcus horikoshii CutA1, which has a denaturation temperature of nearly 150 °c. *FEBS J.* **279,** 78–90 (2012).
243. Huang, C.-W. *et al.* The prognostic values of EGFR expression and KRAS mutation in patients with synchronous or metachronous metastatic colorectal cancer. *BMC Cancer* **13,** 599 (2013).
244. Rokita, M. *et al.* Overexpression of epidermal growth factor receptor as a prognostic factor in colorectal cancer on the basis of the Allred scoring system. *Onco. Targets. Ther.* **6,** 967–976 (2013).
245. Schymkowitz, J. *et al.* The FoldX web server: An online force field. *Nucleic Acids Res.* **33,** W382-8 (2005).
246. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132 (1982).
247. James, E. *et al.* Structural and Mechanistic Basis for the Inhibition of Escherichia coli RNA

Polymerase by T7 Gp2. *Mol. Cell* **47,** 755–766 (2012).
248. Bae, B. *et al.* Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of 70 domain 1.1. *Proc. Natl. Acad. Sci.* **110,** 19772–19777 (2013).
249. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
250. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41,** 597–600 (2013).
251. Traxlmayr, M. W. *et al.* Directed evolution of stabilized IgG1-Fc scaffolds by application of strong heat shock to libraries displayed on yeast. *Biochim. Biophys. Acta - Proteins Proteomics* **1824,** 542–549 (2012).
252. Cho, Y. K. & Shusta, E. V. Antibody library screens using detergent-solubilized mammalian cell lysates as antigen sources. *Protein Eng. Des. Sel.* **23,** 567–577 (2010).
253. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: Paired-end assembler for illumina sequences. *BMC Bioinformatics* **13,** 1–7 (2012).
254. Case, B. A., Kruziki, M. A., Stern, L. A. & Hackel, B. J. Evaluation of affibody charge modification identified by synthetic consensus design in molecular PET imaging of epidermal growth factor receptor. *Mol. Syst. Des. Eng.* **3,** 171–182 (2018).
255. Chen, T. F., de Picciotto, S., Hackel, B. J. & Wittrup, K. D. Chapter Fourteen – Engineering Fibronectin-Based Binding Proteins by Yeast Surface Display. *Methods Enzymol.* **523,** 303–326 (2013).
256. Firth, A. E. & Patrick, W. M. GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Res.* **36,** 281–285 (2008).
257. Ackerman, M. *et al.* Highly avid magnetic bead capture: An efficient selection method for de novo protein engineering utilizing yeast surface display. *Biotechnol. Prog.* **25,** 774–783 (2009).
258. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74,** 5463–5467 (1977).
259. Nielsen, C. H. *et al.* PET Imaging of Tumor Neovascularization in a Transgenic Mouse Model with a Novel 64Cu-DOTA-Knottin Peptide. *Cancer Res.* **70,** 9022–9030 (2010).
260. Fani, M., André, J. P. & Maecke, H. R. 68Ga-PET: a powerful generator-based alternative to cyclotron-based PET radiopharmaceuticals. *Contrast Media Mol. Imaging* **3,** 53–63 (2008).
261. Woldring, D. R., Holec, P. V., Stern, L. A., Du, Y. & Hackel, B. J. A Gradient of Sitewise Diversity Promotes Evolutionary Fitness for Binder Discovery in a Three-Helix Bundle Protein Scaffold. *Biochemistry* **56,** 1656–1671 (2017).
262. Hackel, B. J., Ackerman, M. E., Howland, S. W. & Wittrup, K. D. Stability and CDR Composition Biases Enrich Binder Functionality Landscapes. *J. Mol. Biol.* **401,** 84–96 (2010).